Accurate Metacognition for Visual Sensory Memory Representations

Annelinde R. E. Vandenbroucke, Ilja G. Sligte, Adam B. Barrett, Anil K. Seth, Johannes J. Fahrenfort and Victor A. F. Lamme

Psychological Science published online 18 February 2014
DOI: 10.1177/0956797613516146

Psychological Science OnlineFirst, published on February 18, 2014 as doi:10.1177/0956797613516146

Research Article

Psychological Science
1–13

DOI: 10.1177/0956797613516146
pss.sagepub.com

Accurate Metacognition for Visual Sensory Memory Representations

Annelinde R. E. Vandenbroucke[1,2], Ilja G. Sligte[1,2,3], Adam B. Barrett[4,5], Anil K. Seth[4,5], Johannes J. Fahrenfort[6], and Victor A. F. Lamme[1,2]

[1]Cognitive Neuroscience Group, Department of Psychology, University of Amsterdam; [2]Amsterdam Brain and Cognition Center, University of Amsterdam; [3]Visual Experience Lab, Department of Psychology, University of Birmingham; [4]Sackler Centre for Consciousness Science, University of Sussex; [5]Department of Informatics, University of Sussex; and [6]Department of Psychology, Utrecht University

Abstract

The capacity to attend to multiple objects in the visual field is limited. However, introspectively, people feel that they see the whole visual world at once. Some scholars suggest that this introspective feeling is based on short-lived sensory memory representations, whereas others argue that the feeling of seeing more than can be attended to is illusory. Here, we investigated this phenomenon by combining objective memory performance with subjective confidence ratings during a change-detection task. This allowed us to compute a measure of metacognition—the degree of knowledge that subjects have about the correctness of their decisions—for different stages of memory. We show that subjects store more objects in sensory memory than they can attend to but, at the same time, have similar metacognition for sensory memory and working memory representations. This suggests that these subjective impressions are not an illusion but accurate reflections of the richness of visual perception.

Keywords

attention, visual short-term memory, subjective, visual memory, consciousness, decision making

Received 3/26/13; Revision accepted 11/9/13

In day-to-day life, people subjectively experience everything in their visual field as a rich and integrated whole. When asked about a scene that has disappeared from view, however, they can report only about the few items they happened to attend to. This dissociation between people's rich experience and limited attentional capacities remains poorly understood. The introspective feeling of rich perception has been supported by partial-report studies (Landman, Spekreijse, & Lamme, 2003; Sligte, Scholte, & Lamme, 2008; Sperling, 1960) showing that for a brief moment after disappearance of a visual display, a cue can guide subjects to retrieve much more information from the display than they can when no cue is given. It thus seems that a lot of information is available for a short period after stimulus offset, but this information quickly decays over time.

The existence of this temporary high-capacity memory store has been taken to suggest that conscious experience is not limited to what people can report about. Instead, their limited attentional capacities restrict unattended information from being made robust and available for report and for cognitive manipulations (Block, 2007, 2011; Lamme, 2006, 2010). Other researchers, however, argue that unattended items are never consciously processed, and attention is necessary to have a visual experience (Cohen & Dennett, 2011; Kouider, de Gardelle, Sackur, & Dupoux, 2010). According to this view, the introspective feeling of seeing more than can be attended to is illusory, and high-capacity performance in partial-report experiments is based on implicit or unconscious information (Lau & Rosenthal, 2011; Rahnev et al., 2011). In the present study, we investigated whether the subjective experience of perceiving more than can be attended

Corresponding Author:
Annelinde R. E. Vandenbroucke, University of Amsterdam, Department of Psychology, Roetersstraat 15, Amsterdam 1018WB, The Netherlands
E-mail: vandenbroucke.work@gmail.com

to is real or illusory by combining objective with subjective ratings in a partial-report experiment, thereby measuring the level of metacognition for unattended memory representations (Fleming, Weil, Nagy, Dolan, & Rees, 2010; Kanai, Walsh, & Tseng, 2010).

Previous studies have shown that when subjects attend away from a stimulus location, they adopt a liberal response bias for targets in the unattended location, tending to report that a stimulus is present. In contrast, when subjects attend to a stimulus location, they adopt a conservative response bias and less often report that a stimulus is present. In addition, the confidence ratings accompanying perceptual decisions are higher for unattended than for attended stimuli (Rahnev, Bahdo, de Lange, & Lau, 2012; Rahnev et al., 2011; Rahnev, Maniscalco, Luber, Lau, & Lisanby, 2012; Wilimzig, Tsuchyia, Fahle, Einhäuser, & Koch, 2008). Although this seems counterintuitive, it can be explained within the framework of signal detection theory (see Fig. 1; Macmillan & Creelman, 2005). It has therefore been suggested that people's subjectively rich perception is inflated, and actually very little is seen outside the focus of attention (Rahnev et al., 2011).

A crucial point that has been overlooked, however, is whether subjective confidence ratings coincide with

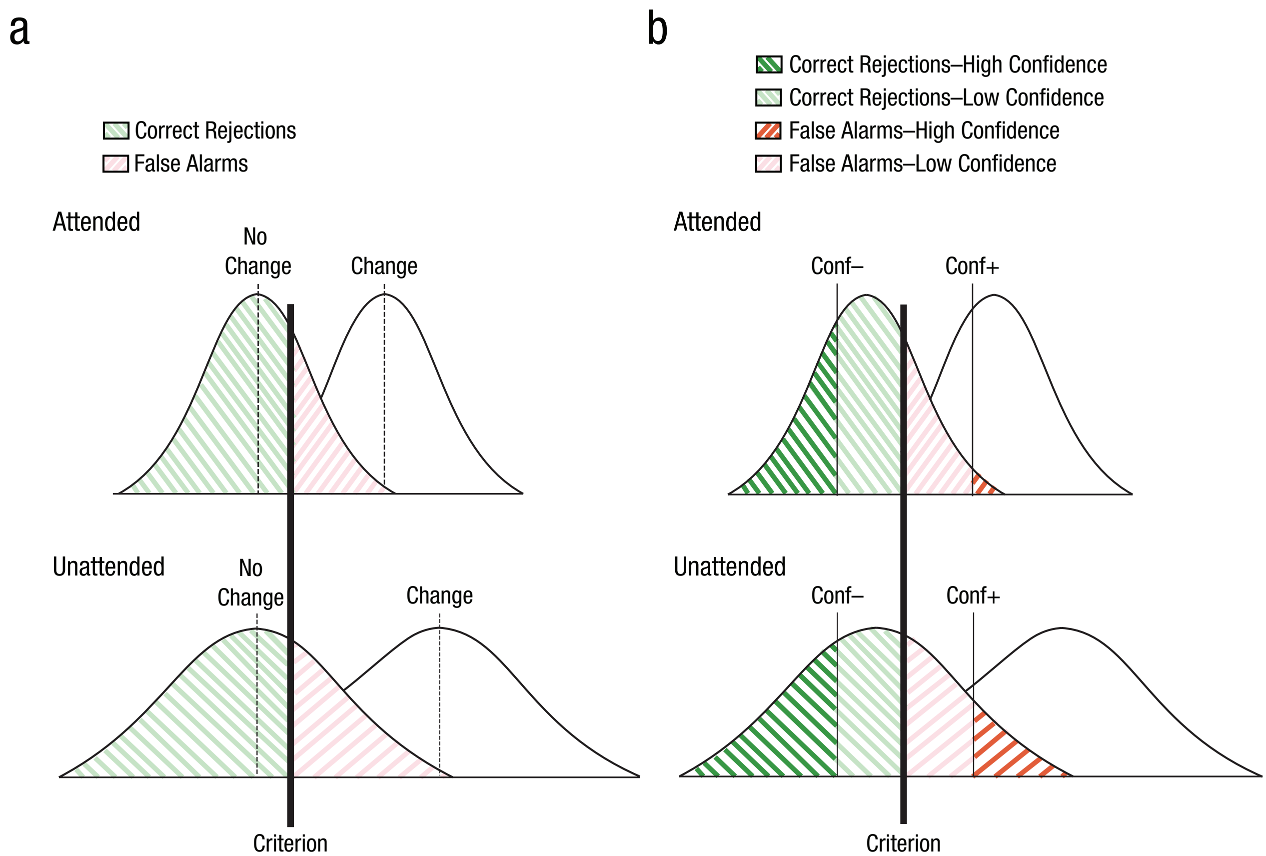

Fig. 1. Hypothetical signal detection diagrams under conditions in which subjects attend (upper panels) and do not attend (lower panels) to stimulus location. Hypothetical probability densities of signal strength induced by two stimuli in a change-detection paradigm are shown in (a). In this case, a stimulus that does not change is presented. The probability densities of change and no change are normally distributed and are a certain distance apart. The distance between the peaks divided by their spread (in standard-deviation units, also known as d') signals the discriminability between the two alternatives. The decision criterion (thick vertical line) is the threshold above which a subject responds that the stimulus has changed in a particular trial. When the decision criterion lies exactly between the two peaks, it will result in the same number of "no-change" and "change" responses (assuming equal variance). When the decision criterion is shifted toward either peak, there will be a corresponding response bias; in this figure, the response bias is negative (the criterion is closer to the left peak than to the right peak), meaning that subjects respond "change" more often, which results in more hits but also more false alarms. When a stimulus is unattended (lower panel), there is more variance in the signal than when the stimulus is attended (upper panel). This creates a wider signal distribution, but when the decision criterion remains the same, this will result in more "change" responses (again, assuming equal variance) and thus a higher false alarm rate (Rahnev et al., 2011). In (b), a hypothetical effect of attention on confidence ratings is shown (Rahnev et al., 2011). When the signal is above the greater confidence (conf+) threshold or below the lower confidence (conf–) threshold, subjects will give high confidence ratings; when the signal is between the conf– and conf+ thresholds, subjects will give low confidence ratings. Conf– is the threshold for the subject having high confidence that the stimulus did not change, and conf+ is the threshold for the subject having high confidence that the stimulus changed when in fact it did not. Again, in the absence of attention, there is more variance in the signal (compared with the upper panel). This results in a wider distribution, and hence a larger number of high confidence ratings, assuming that the confidence criterion itself does not change between attended and unattended stimuli. (Characteristics are described here using a change-detection paradigm, because we used this paradigm in Experiment 1. To translate this diagram to classical signal detection theory, replace "no change" with "stimulus absent" and "change" with "stimulus present.")

objective performance. When correct responses are accompanied by high confidence and incorrect responses by low confidence, subjects have good knowledge about the correctness of their perceptual decisions, or high metacognitive performance (Fleming et al., 2010; Metcalfe & Shimamura, 1994). Alternatively, when there is no relationship between confidence ratings and the correctness of perceptual decisions, metacognition is low. Investigating subjective and objective ratings in isolation reveals the characteristics of decision criteria and confidence, but looking at metacognitive performance reveals whether subjects base objective decisions on explicit knowledge.

To investigate metacognition for unattended visual representations, we combined a partial-report change-detection paradigm with subjective confidence ratings. By using a partial-report paradigm, one can distinguish between visual sensory memory (Neisser, 1967; Sperling, 1960) and visual working memory (Luck & Vogel, 1997). Visual sensory memory is a high-capacity memory store in which information is maintained in a fragile format for a short period of time (Sligte et al., 2008). Its high capacity is measured by providing subjects with a cue after memory-display offset but before test-display onset. Visual sensory memory can be divided into iconic memory and fragile visual short-term memory. Iconic memory is a high-capacity, short-lived store that is partially dependent on afterimages. Fragile memory, in contrast, can last up to 4 s (Sligte et al., 2008) and is supported by cortical processing (Sligte, Scholte, & Lamme, 2009), but it is fragile because it is overwritten by a new display containing similar items (Makovski, Sussman, & Jiang, 2008; Pinto, Sligte, Shapiro, & Lamme, 2013). This fragility is in contrast with working memory, a low-capacity, long-lived storage that is not overwritten by new displays (Baddeley & Hitch, 1986; Durstewitz, Seamans, & Sejnowski, 2000; Pinto et al., 2013). Crucially, working memory capacity depends on attention (Awh, Vogel, & Oh, 2006; Chun, 2011), whereas fragile-memory capacity is hardly reduced when attention is diverted during memory encoding (Vandenbroucke, Sligte, & Lamme, 2011).

Because working memory contains information we can manipulate and report about, it is thought to reflect conscious, explicit processing (Baars & Franklin, 2003; Lamme, 2006). Therefore, we compared metacognition for working memory to metacognition for sensory memory: If representations in sensory memory reflect implicit information processing, metacognitive performance should be lower for sensory memory than for working memory. If, however, the information stored in sensory memory is as explicit as information stored in working memory, metacognitive performance should be equal.

Measures of metacognition are notoriously subject to biases and confounds (Galvin, Podd, Drga, & Whitmore, 2003), which we were careful to control for. First, to ensure that differences in subjective scores cannot be ascribed to differences in objective performance (Lau & Passingham, 2006), we adopted a staircase procedure in which objective performance in all conditions was kept at 75% by varying the number of items to remember. This allowed us to measure capacity differences between sensory and working memory while keeping task difficulty the same. Second, we applied a recently introduced measure, meta-*d*′-balance (Barrett, Dienes, & Seth, 2013; Maniscalco & Lau, 2011), that complements standard signal detection analysis to ensure that metacognitive scores were not confounded by variation in objective or subjective decision criteria.

In Experiment 1, subjects performed a change-detection task on stimulus orientation. We found that in the case of iconic memory, metacognition was similar to metacognition for working memory, and in the case of fragile memory, it was even higher. In addition, subjects adopted a more liberal response bias for sensory memory than for working memory; that is, in sensory-memory conditions, subjects reported perceiving a change more often than in the working-memory condition. This matches the response bias found by Rahnev et al. (2011) for unattended versus attended stages, respectively, and confirms our previous findings that fragile memory and iconic memory represent unattended stages of memory processing, whereas working memory reflects attended processing (Vandenbroucke et al., 2011). However, the differences in response bias among conditions, combined with the high hit and false alarm rates observed in this experiment, might have influenced metacognition scores (Barrett et al., 2013). We therefore conducted a second experiment in which we equalized the objective response bias. Subjects performed a discrimination task instead of a detection task: Stimuli always changed orientation, and subjects indicated whether they perceived a clockwise or counterclockwise change. As this was not a detection task, in which the stimulus had to pass a certain threshold to be reported as seen, we expected the response bias to be close to 0. We found that metacognition was now equal for sensory memory and working memory, and because the response bias was equal for all three memory conditions, the comparison of metacognition among iconic memory, fragile memory, and working memory was fully warranted.

Experiment 1

Method

Subjects. Twenty-five students (9 men, 16 women; mean age = 21.5 years, *SD* = 1.9) of the University of Amsterdam participated in this experiment. All reported that they had normal or corrected-to-normal vision.

Subjects gave written informed consent before the experiment, which was approved by the local ethics committee.

Task, procedure, and stimuli. Stimulus displays consisted of white rectangles (1.55° × 0.40° of visual angle) on a black background; the white rectangles were located randomly in the 36 squares of a 6 × 6 grid (12.24° × 12.24°; see Fig. 2). The rectangles had four possible orientations: horizontal, vertical, 45° to the horizontal, or 135° to the horizontal. The cue consisted of four triangles (short sides = 0.23°) positioned in each corner of one of the placeholder (Fig. 2).

In all conditions, a memory array appeared for 250 ms at the start of each trial. In the sensory-memory conditions (i.e., the iconic-memory and fragile-memory

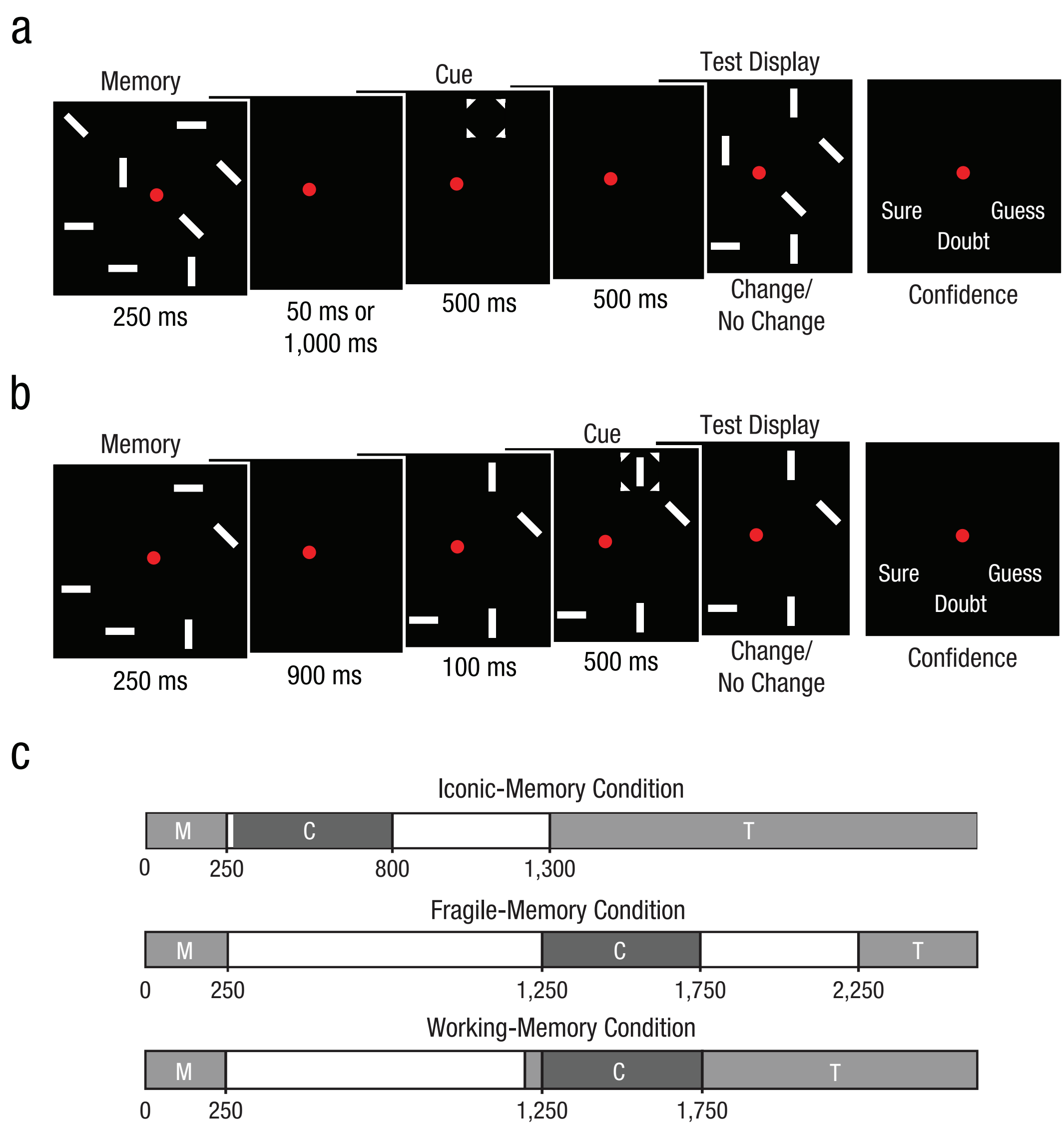

Fig. 2. Example trial sequence and cue timings for the three conditions in Experiment 1. In the iconic-memory and fragile-memory conditions (a), a memory display containing a number of white rectangles around a central fixation point appeared at the start of each trial. This was followed by a blank interstimulus interval (ISI) of either 50 ms (iconic-memory condition) or 1,000 ms (fragile-memory condition). A cue then appeared briefly, followed by a second ISI. Subjects were then shown a test display, and they had to indicate whether it was the same as the memory display (no change) or whether one of the rectangles was in a different orientation (change). Subjects were then asked to judge their confidence in their perceptual decision by choosing one of three answers: sure (positive their answer was correct), doubt (reasonably sure their answer was correct), or guess (not sure at all). In the working-memory condition (b), the ISI following the memory display was followed by a second presentation of the memory display. The cue was then presented while the rectangles remained on-screen. After the cue disappeared, subjects had to make a change/no-change decision and give confidence judgments as in the other trial types. The schematic (c) shows a comparison of the timings of the memory display (M), cue (C), and test display (T).

conditions), an interstimulus interval (ISI) of either 50 ms (iconic-memory condition) or 1,000 ms (fragile-memory condition) occurred after offset of the memory array. A cue then appeared for 500 ms, followed by a 500-ms ISI. Presenting the cue after offset of the memory array but before onset of the test display allowed subjects to retrieve the information that was maintained before interference of a new display and resulted in a larger number of rectangle orientations that could be remembered. In the working-memory condition, the ISI following the memory display was 900 ms, and it was followed by a 100-ms representation of the memory display. The cue was then presented for 500 ms while the rectangles remained on-screen. Because the onset of the test array erased all sensory memory traces, this procedure ensured that only working memory was measured, although subjects still knew which item was the only item that might have changed.

In all conditions, at presentation of the test display, subjects indicated whether they had perceived a change between the memory and test display. A change occurred on 50% of trials and always consisted of one of the rectangles rotating 90°. The cue was always valid, that is, it indicated which item would change (sensory-memory conditions) or could have changed (working-memory condition). After this objective rating (change or no change), subjects were asked to judge their confidence in their perceptual decision by choosing one of three answers: sure (positive their answer was correct), doubt (reasonably sure their answer was correct), or guess (not sure at all). Subjects were encouraged to use all three options throughout the experiment.

Before testing began, subjects received a training of 60 trials (because of a technical mistake, 5 subjects received 10–30 additional training trials). Iconic-memory, fragile-memory, and working-memory trials were randomly intermixed (20 trials each), and subjects were not informed about trial type. During training, the displays contained six randomly placed rectangles. Subjects received immediate feedback on the correctness of their objective response (confidence judgments were not elicited during training).

Previous studies have shown that capacity for iconic memory and fragile memory is much higher than capacity for working memory (Sligte et al., 2008; Vandenbroucke et al., 2011), and there are large individual differences. Because objective performance can influence subjective performance (Lau & Passingham, 2006), objective performance for all memory conditions was kept at 75%. On the basis of the results of the training block, we computed the initial number of rectangles that would be used in the experimental blocks for iconic memory, fragile memory, and working memory separately. When performance was 75% correct for a condition during training (six objects; chance level of 50%), the initial set size for that condition in the experimental block was six objects. For each 15% that subjects scored below 75%, one rectangle was subtracted from the display at the beginning of the experimental trials. For each 15% that subjects scored above 75% correct, a rectangle was added to the displays at the start of the experimental trials.

After training, subjects performed an experimental block of 366 trials (of which the first 6 were not analyzed) in which the immediate feedback (change/no-change decision) was eliminated, and subjects additionally provided confidence judgments (2 subjects received 306 trials each). To keep objective performance the same on all conditions and constant over the course of the block, we calculated the percentage of correct responses on every 4 trials (per condition), and a rectangle was added to or removed from the displays when performance was higher or lower than 75%. This resulted in an average performance of 75% for each condition, but a different capacity score, which was defined by the number of rectangles present in the display at the end of the experimental block (Fig. 3b).

Results

To evaluate objective performance, we calculated sensitivity as Type I d' (z-scored hit rate – z-scored false alarm rate; Green & Swets, 1966). Hits were classified as correctly reported changes, and false alarms were classified as incorrectly reported changes. We excluded 1 subject from the analyses, because as the result of a technical mistake, this subject performed the task twice in a row, and performance dropped throughout the second run. Figure 3a shows that—as intended—d' for fragile memory and working memory did not significantly differ, $t(23) = -0.9$, $p = .380$, but d' for iconic memory was slightly higher, especially compared with fragile memory, $t(23) = 3.4$, $p = .002$; there was also a main effect of memory conditions, $F(2, 46) = 4.4$, $p = .018$. Exploring the performance level for each condition showed that a few subjects kept on improving their score for iconic memory throughout the experiment. This suggests that the iconic-memory condition was easier than the fragile-memory and working-memory conditions, and therefore, d' over the whole experiment was somewhat higher. The manipulation of keeping performance at 75% resulted in a different number of rectangles in the displays at the end of the experiment for each of the three conditions (Fig. 3b), $F(2, 46) = 29.5$, $p < .001$: Iconic-memory capacity was higher than fragile-memory capacity, $t(23) = 2.4$, $p = .026$, which was higher than working memory capacity, $t(23) = 6.5$, $p < .001$.

To investigate response bias, we computed c using the following formula: $-0.5 \times$ (z-scored hit rate + z-scored false alarm rate; Green & Swets, 1966). Response bias was negative for both sensory-memory conditions, which shows that there was a tendency to respond "change" more often than "no change," whereas in the working-memory condition, the opposite occurred (Fig. 3c), $F(2, 46) = 132.7$, $p < .001$. Thus, for sensory memory, a more liberal response bias was adopted, whereas for working memory, the criterion was more conservative.

Before we assessed metacognitive performance, we analyzed whether confidence ratings themselves differed among memory conditions. We calculated mean confidence ratings by multiplying the proportion of given confidence ratings by their rank number ("sure" = 3, "doubt" = 2, "guess" = 1). There were no significant differences among the three memory types, $F(2, 46) = 3.3$, $p = .059$, Greenhouse-Geisser corrected, although there was a trend toward a difference that was probably caused by the lower mean confidence rating for fragile memory (2.48) versus

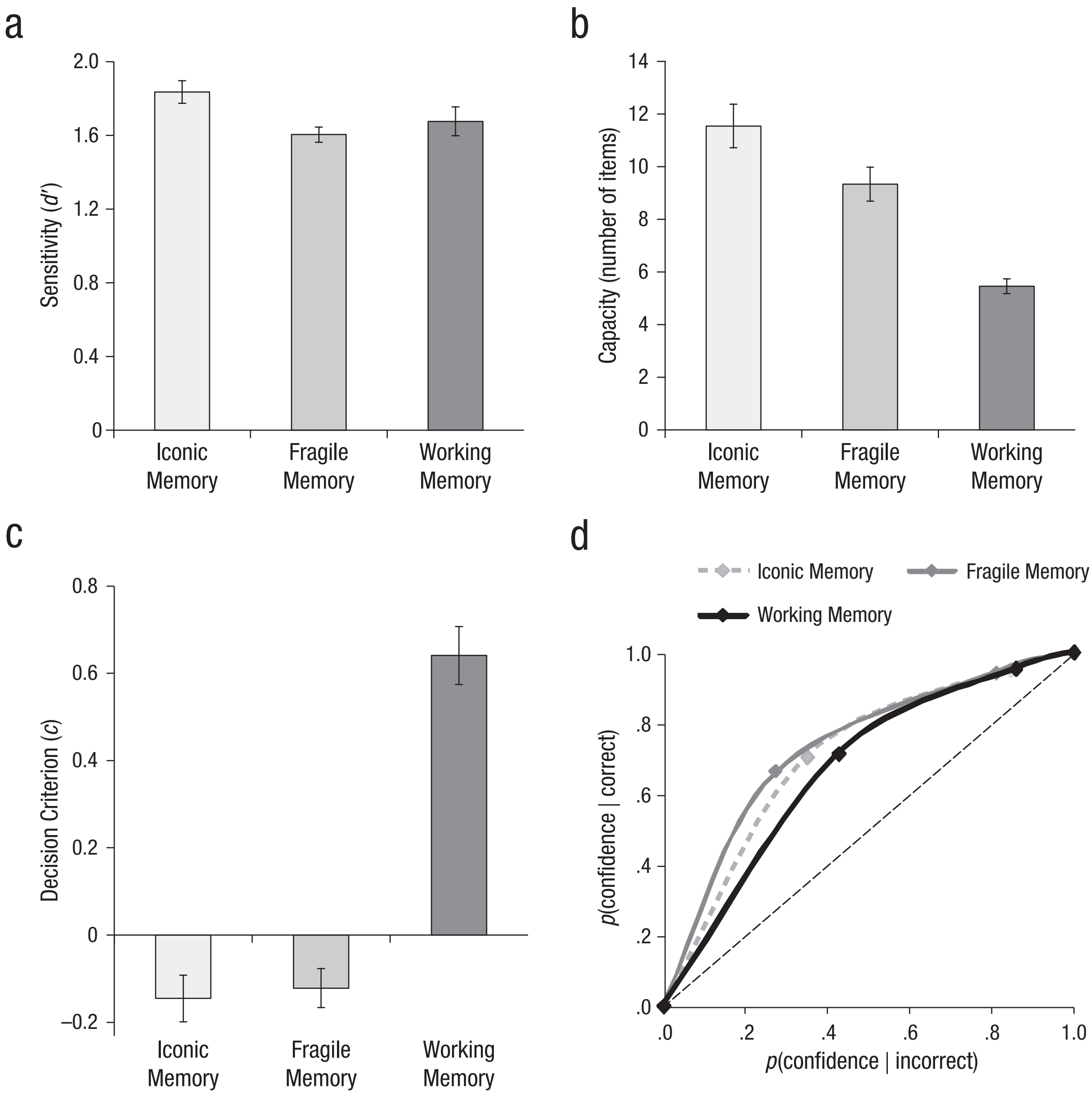

Fig. 3. Results from Experiment 1. Mean sensitivity (z-scored hit rate − z-scored false alarm rate, represented as d') is shown in (a) as a function of condition. The graph in (b) shows the mean number of rectangles in each condition that participants were able to hold in memory. Mean response bias (expressed in c using standard-deviation units) for each condition is shown in (c). To calculate c, we used the following formula: $-0.5 \times$ (z-scored hit rate + z-scored false alarm rate). Error bars in (a) through (c) show standard errors of the mean. The graph in (d) illustrates receiver operating characteristic curves that show the ability to discriminate between incorrect and correct responses for each level of confidence. For each condition, the points plotted are the rates of Type II hits (confident when correct) and false alarms (confident when incorrect) when, from left to right, no responses were classified as confident, only "sure" responses were classified as confident, both "sure" and "doubt" responses were classified as confident, and all responses were classified as confident.

iconic memory and working memory (2.55 and 2.56, respectively).

The level of metacognition was established by analyzing the Type II receiver operating characteristic (ROC) curve (Fig. 3d). The ROC curve plots the cumulative probabilities of the confidence ratings for correct versus incorrect responses (Macmillan & Creelman 2005). The area under the (ROC) curve (AUC; deviation from the diagonal) then provides a measure of the ability to link confidence to perceptual performance (Fleming et al., 2010; Galvin et al., 2003). The AUC for the three conditions differed, $F(2, 46) = 4.5$, $p = .013$, and this effect was mainly driven by the fact that the AUC was larger for fragile memory than for working memory, $t(23) = 3.3$, $p = .003$. This suggests that metacognition for iconic memory and working memory are similar, $t(23) = 1.5$, $p = .14$, and metacognition for fragile memory might even be higher.

Although one might conclude from this analysis that metacognition for sensory memory is similar to or even higher than metacognition for working memory, it has been shown that Type I response bias can influence metacognition scores based on the AUC (Barrett et al., 2013; Maniscalco & Lau, 2011). Therefore, we also calculated meta-d', a newly developed measure of metacognition that is less sensitive to response-bias variation (Maniscalco & Lau, 2011; Barrett et al., 2013). This measure is defined as the Type I d' that would have led to the observed Type II (confidence) data had the subject's response and confidence judgment both followed the standard signal-detection-theory model. Several algorithms have been proposed for computing meta-d', each of which deals in a different way with the fact that metacognition can actually be different for positive and negative Type I responses. We utilized the recently developed *meta-d′-balance* statistic (Barrett et al., 2013), which weighs values for positive and negative responses proportionally to the respective frequencies of positive and negative responses, and admits a unique solution given the Type I and Type II data. When we included all subjects, the results for these analyses were similar to those obtained from the ROC analysis (see Table 1), which shows that fragile memory had a higher metacognition score than iconic memory and trended toward a higher metacognition score for working memory. Taken together, these analyses suggest that metacognition for sensory memory is similar to or even higher than metacognition for working memory.

Although meta-d'-balance is robust to variation in response bias, it can still deliver unstable estimates for extreme hit and false alarm rates (< .05 or > .95; see Barrett et al., 2013). When we excluded subjects with estimated responses in these ranges, only 3 subjects remained in the working-memory condition. Therefore, in order to rigorously compare metacognition for working memory and sensory memory, we conducted a second experiment that was designed to maintain a response bias closer to 0 in all conditions.

Table 1. Meta-d'-Balance Results From Experiment 1 for All Subjects and for a Subset of Subjects

	All subjects included (narrow exclusion criteria)			Only stable subjects included (wide exclusion criteria)		
Condition	Meta-d'-balance	N	Comparison condition	Meta-d'-balance	n	Comparison condition
Iconic memory	1.13 (1.06)	24	Fragile memory: $p = .04$ Working memory: $p = .78$	0.64 (0.11)	13	Fragile memory: $p = .81$ Working memory: $p = .72$
Fragile memory	2.04 (1.84)	24	Iconic memory: $p = .04$ Working memory: $p = .08$	0.70 (0.19)	9	Iconic memory: $p = .81$ Working memory: $p = .80$
Working memory	1.23 (1.30)	24	Iconic memory: $p = .78$ Fragile memory: $p = .08$	0.53 (0.40)	3	Iconic memory: $p = .80$ Fragile memory: $p = .72$

Note: The two sets of criteria for subject exclusions were suggested by Barrett, Dienes, and Seth (2013). Narrow exclusion criteria include all subjects whose estimated hit and false alarm rates were all greater than 0 and less than 1, and wide exclusion criteria include only subjects whose estimated hit and false alarm rates were greater than .05 and lower than .95. Standard deviations are given in parentheses. Comparisons between conditions were made using unpaired t tests.

Experiment 2

From Experiment 1, one might conclude that metacognition for sensory memory is equal to (iconic-memory condition) or higher (fragile-memory condition) than metacognition for working memory. However, variation in the response bias might have affected metacognition scores by leading to extreme false alarm and hit rates (Barrett et al., 2013; Maniscalco & Lau, 2011). To better compare metacognition for the three memory types, in Experiment 2, we equated decision criteria by substituting the change-detection task with a change-identification task. The rectangles were replaced by arrows, and the cued arrow always changed orientation between memory and test display. The subjects' task was now to indicate whether the change in orientation was clockwise or counterclockwise. Because this was not a detection task, in which the stimulus had to pass a certain threshold to be reported as seen, there was no reason to expect a bias between responding clockwise or counterclockwise; the response bias should be closer to 0 and be roughly equal for attended (working memory) and unattended (iconic memory, fragile memory) representations.

Method

Subjects. Twenty-four students (3 men, 21 women; mean age = 22.0 years, $SD = 0.6$) of the University of Amsterdam participated in this experiment. All reported that they had normal or corrected-to-normal vision. Subjects gave written informed consent before the experiment, which was approved by the local ethics committee.

Task, procedure, and stimuli. The task and procedure were the same as those in Experiment 1, except now subjects indicated whether a cued arrow had changed orientation clockwise or counterclockwise. The arrows (1.20° × 0.63°) were oriented up, down, left, or right.

Results

The manipulation of keeping objective performance similar was successful, as reflected by d' being not significantly different among the three memory conditions (Fig. 4a), $F(2, 46) = 1.6$, $p = .205$ (hits were classified as correctly reported counterclockwise changes, and false alarms were classified as incorrectly reported counterclockwise changes). As in Experiment 1, capacity scores differed among conditions (Fig. 4b), $F(1.4, 32.6) = 7.2$, $p = .006$, Greenhouse-Geisser corrected. However, there was no difference between iconic-memory and fragile-memory capacity, $t(23) = -0.5$, $p = .634$, but only a difference between each of these two conditions and working memory capacity—iconic memory vs. working memory: $t(23) = 4.9$, $p < .001$; fragile memory vs. working memory: $t(23) = 3.0$, $p = .007$. As expected, response biases were now also similar for each memory condition (Fig. 4c), $F(2, 46) = 0.7$, $p = .513$, and therefore, the comparison between metacognition scores using AUC can be fully justified (Barrett et al., 2013). Also, there were no significant differences across conditions in mean confidence ratings (iconic memory: 2.38; fragile memory: 2.37, working memory: 2.39), $F(2, 46) = 0.4$, $p = .674$.

Metacognition scores, as calculated by the AUC of the ROC (Fig. 4d), were just significantly different for the three memory conditions, $F(2, 46) = 3.2$, $p = .049$. This difference was driven by the fact that metacognition was lower for iconic memory than for working memory, $t(23) = -2.3$, $p = .030$. Metacognition scores for fragile memory and working memory, however, were now not significantly different, $t(23) = -0.6$, $p = .531$, which suggests that when controlling for objective sensitivity and decision criteria, fragile memory and working memory are equally based on explicit processing. (See the Supplemental Material available online for a Bayes factor analysis that indicates the likelihood of metacognition for fragile memory and working memory being the same.)

Using meta-d'-balance, we found similar results indicating that metacognition for fragile memory and working memory were similar, whereas metacognition for iconic memory tended to be lower (Table 2). In addition, when using a wide exclusion criterion, fewer subjects were excluded than in Experiment 1, which confirms that there were fewer extremely response-biased subjects in Experiment 2.

Discussion

In the present experiments, we measured metacognitive performance on a partial-report change-detection task to investigate whether early (iconic memory) and late (fragile memory) sensory memory can be accessed explicitly. We compared metacognition for sensory memory to metacognition for working memory, which is low in capacity, explicit, and attention dependent. At equal objective performance (d'), iconic-memory and fragile-memory capacity were higher than working memory capacity. At the same time, metacognition for fragile memory was higher than (Experiment 1) or equal to (Experiment 2) metacognition for working memory. This suggests that the higher capacity of fragile memory is not based on implicit, unconscious information but reflects explicit and possibly conscious information processing.

Previous research has shown that iconic memory and fragile memory reflect unattended processing. When attention is diverted during encoding of a memory array,

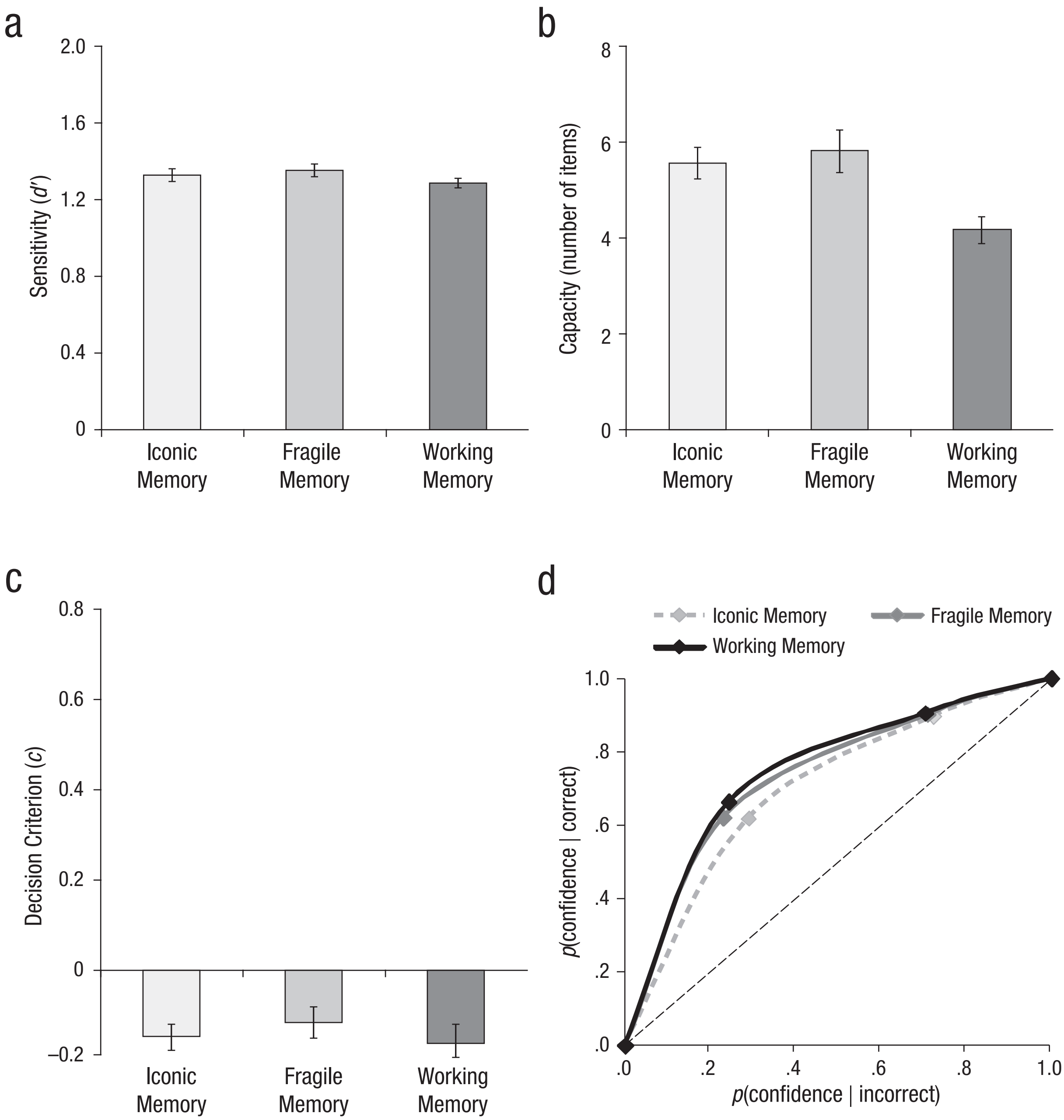

Fig. 4. Results from Experiment 2. Mean sensitivity (*z*-scored hit rate – *z*-scored false alarm rate, represented as d') is shown in (a) as a function of condition. The graph in (b) shows the mean number of rectangles in each condition that participants were able to hold in memory. Mean response bias (expressed in *c* using standard-deviation units) for each condition is shown in (c). To calculate *c*, we used the following formula: –0.5 × (*z*-scored hit rate + *z*-scored false alarm rate). Error bars in (a) through (c) show standard errors of the mean. The graph in (d) illustrates receiver operating characteristic curves that show the ability to discriminate between incorrect and correct responses for each level of confidence. For each condition, the points plotted are the rates of Type II hits (confident when correct) and false alarms (confident when incorrect) when, from left to right, no responses were classified as confident, only "sure" responses were classified as confident, both "sure" and "doubt" responses were classified as confident, and all responses were classified as confident.

working memory capacity suffers, whereas fragile-memory capacity remains relatively intact (Vandenbroucke et al., 2011). This suggests that fragile-memory representations are formed independently of focused attention. Moreover, when subjects receive two serial retro-cues instead of one (the second cue pointing toward a different location), performance on the second cued item stays high (Landman et al., 2003). This suggests that even though

Table 2. Meta-*d′*-Balance Results From Experiment 2 for All Subjects and for a Subset of Subjects

	All subjects included (narrow exclusion criteria)			Only stable subjects included (wide exclusion criteria)		
Condition	Meta-*d′*-balance	*N*	Comparison condition	Meta-*d′*-balance	*n*	Comparison condition
Iconic memory	1.05 (1.48)	24	Fragile memory: $p = .35$ Working memory: $p = .15$	0.71 (0.47)	20	Fragile memory: $p = .33$ Working memory: $p = .31$
Fragile memory	1.42 (1.27)	24	Iconic memory: $p = .35$ Working memory: $p = .50$	0.86 (0.42)	14	Iconic memory: $p = .33$ Working memory: $p = .93$
Working memory	1.71 (1.64)	24	Iconic memory: $p = .15$ Fragile memory: $p = .50$	0.88 (0.52)	18	Iconic memory: $p = .31$ Fragile memory: $p = .93$

Note: The two sets of criteria for subject exclusions were suggested by Barrett, Dienes, and Seth (2013). Narrow exclusion criteria include all subjects whose estimated hit and false alarm rates were all greater than 0 and less than 1, and wide exclusion criteria include only subjects whose estimated hit and false alarm rates were greater than .05 and lower than .95. Standard deviations are given in parentheses. Comparisons between conditions were made using unpaired *t* tests.

attention is directed toward one sensory memory representation, the other sensory memory representations stay available for report.

In the current study, subjects had equal metacognitive performance for sensory memory and working memory representations. As attention is not necessary for sensory memory to be formed and maintained (Landman et al., 2003; Vandenbroucke et al., 2011), we conclude that unattended, sensory memory items are a meaningful part of visual experience. However, we only measured metacognitive performance for the cued, and not for the uncued, iconic-memory and fragile-memory representations. It could be argued that all properties that are measured in this paradigm are only about the cued item: Subjects report seeing a change in one object and report metacognition for that single object. According to this view, the paradigm says nothing about the uncued items. However, because subjects do not know in advance which item will be cued, performance on the cued item represents performance on any other item in the display. Following the logic of the original Sperling (1960) experiment, one can thus derive memory capacity from performance on the cued item multiplied by the number of items present in the display. This same argument applies to the observation of equal metacognitive performance for sensory memory and working memory: Metacognitive performance on the cued item represents metacognition for any item in the display. Thus, there is accurate metacognition for all items represented in high-capacity sensory memory.

A challenge when measuring metacognition is that subjects can give confidence ratings only about reported items. One could thus argue that access to an item's representation is necessary for metacognition to arise. Possibly, after cuing one item, all item representations have changed, and the uncued items are lost. On the basis of the present findings, we cannot draw any conclusions about the level of metacognition for the uncued items after one item has been cued. Previous research has shown that uncued items stay available for report about a change (Landman et al., 2003), which suggests that metacognitive performance for these items stays intact as well. However, this remains to be empirically examined. What can be concluded on the basis of the current experiments is that the information required to support high metacognition on the entire capacity during iconic memory and fragile memory must have been present up to the point of cue presentation and, in that sense, was "rich" during the entire interval.

In addition to previous work showing that sensory memory is not dependent on focused attention, we found that response biases for iconic memory and fragile memory were liberal compared with such biases for working memory (Experiment 1); subjects more often reported perceiving a change in sensory-memory conditions than in the working-memory condition. This finding matches earlier findings that the response bias for unattended representations is more liberal than for attended representations (Rahnev et al., 2011, Rahnev, Maniscalco, et al., 2012), thereby further supporting the claim that sensory memory and working memory capacity reflect unattended versus attended stages in visual short-term memory (Block, 2011; Lamme, 2006, 2010; Sligte et al., 2008; Sligte, Wokke, Tesselaar, Scholte, & Lamme, 2011; Vandenbroucke et al., 2011). However, because response bias might influence metacognitive scores when combined with extreme hit rates and false alarm rates, we conducted a second experiment in which both sensitivity and response bias were equated over conditions.

When sensitivity and response bias were constant across conditions (Experiment 2), metacognition for fragile memory was not significantly different from metacognition for working memory, but metacognition for iconic memory was lower. This suggests a deviation between iconic memory and fragile memory in their dependence on explicit availability. Possibly, the mechanisms underlying iconic memory are partly implicit. This might also explain why for Experiment 1 (detection), iconic-memory capacity was higher than fragile-memory capacity, whereas for Experiment 2 (discrimination), capacity for iconic memory was equal to capacity for fragile memory. Discriminating which type of orientation change occurred might be more complex and dependent on explicit processing compared with simply detecting orientation changes (Clifford, Arnold, & Pearson, 2003). The results of this study therefore suggest that fragile memory has a larger capacity than working memory and at the same time depends on explicit processing just as working memory does, whereas the larger capacity found for iconic memory might partly depend on implicit processing.

Our manipulation of keeping performance the same over memory conditions involved the addition or subtraction of an item in the stimulus displays. This altered the way in which attention was administered to the displays, causing a larger spread in the iconic-memory and fragile-memory conditions—and thus less focused attention—compared with the working-memory condition. In this study, we intended to investigate capacity under conditions of inattention but also under conditions of clear visibility. This is important, because when measuring memory representations, one needs to make sure that a failure to report an item is due to a failure in memory and not due to a failure in perception. This is a different manipulation than that used by Rahnev et al. (2011), in which visibility, and thus ease of detection, was manipulated. The difference between our findings and theirs might be caused by this different manipulation: Although subjects might have similar metacognition for attended and unattended stimuli that are suprathreshold—and thus clearly visible—subjective ratings for stimuli that are near the visibility threshold might be overinterpreted when unattended. This study therefore does not fully resolve the debate around this issue, and further research into the difference between metacognition for memory capacity and stimulus threshold detection should elucidate whether their underlying mechanisms diverge.

The current study is in line with our previous work showing that fragile memory is perceptual in nature (Vandenbroucke, Sligte, Fahrenfort, Ambroziak, & Lamme, 2012): When subjects have to remember illusory triangles (Kanizsa figures) versus unbound control figures, there is a benefit for both fragile memory and working memory. In addition, identification of real-life objects is possible in fragile memory (Sligte, Vandenbroucke, Scholte, & Lamme, 2010). This shows that people not only have increased capacity for simple oriented bars, but also for complex, bound figures, and this suggests that fragile memory reflects integrated information. On top of that, people have similar metacognition for fragile-memory representations. It remains a subject for debate whether metacognition reveals something about the phenomenology of unattended items (Seth, Dienes, Cleeremans, Overgaard, & Pessoa, 2008). Nevertheless, together, these studies show that sensory memory contains higher-order, integrated information, and subjects can subjectively evaluate their access to these items, just as they can evaluate access to working memory items. Therefore, subjective impressions of sensory memory content are no less meaningful than subjective impressions of working memory content.

Author Contributions

A. R. E. Vandenbroucke developed the study concept. A. R. E. Vandenbroucke and I. G. Sligte designed the study. Testing and data collection were performed by A. R. E. Vandenbroucke. A. R. E. Vandenbroucke and A. B. Barrett analyzed and interpreted the data under the supervision of I. G. Sligte and V. A. F. Lamme. A. R. E. Vandenbroucke drafted the manuscript, and I. G. Sligte, A. B. Barrett, A. K. Seth, J. J. Fahrenfort, and V. A. F. Lamme provided critical revisions. All authors approved the final version of the manuscript for submission.

Acknowledgments

We thank Francis Zoet for helping to collect parts of the data. We thank Eric-Jan Wagenmakers for help on the Bayes factor analysis.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

This project was funded by an Advanced Investigator Grant from the European Research Council to V. A. F. Lamme. A. K. Seth and A. B. Barrett were supported by an Engineering and Physical Sciences Research Council Fellowship (EP/G007543/1, awarded to A. K. Seth) and by the Dr. Mortimer and Dame Theresa Sackler Foundation. A. K. Seth was also supported by the European Research Council Collective Experience of Empathic Data Systems project (FP7-ICT-258749), and A. B. Barrett received additional support from Engineering and Physical Sciences Research Council Fellowship EP/L005131/1.

Supplemental Material

Additional supporting information may be found at http://pss.sagepub.com/content/by/supplemental-data

References

Awh, E., Vogel, E., & Oh, S. H. (2006). Interactions between attention and working memory. *Neuroscience*, *139*, 201–208.

Baars, B. J., & Franklin, S. (2003). How conscious experience and working memory interact. *Trends in Cognitive Sciences*, *7*, 166–172.

Baddeley, A. D., & Hitch, G. (1986). *Working memory*. Oxford, England: Clarendon.

Barrett, A. B., Dienes, Z., & Seth, A. K. (2013). Measures of metacognition on signal-detection theoretic models. *Psychological Methods*, *18*, 535–552. doi:10.1037/a0033268

Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral & Brain Sciences*, *30*, 481–498.

Block, N. (2011). Perceptual consciousness overflows cognitive access. *Trends in Cognitive Sciences*, *15*, 567–575.

Chun, M. M. (2011). Visual working memory as visual attention sustained internally over time. *Neuropsychologia*, *49*, 1407–1409.

Clifford, C. W., Arnold, D. H., & Pearson, J. (2003). A paradox of temporal perception revealed by a stimulus oscillating in colour and orientation. *Vision Research*, *43*, 2245–2253.

Cohen, M. A., & Dennett, D. C. (2011). Consciousness cannot be separated from function. *Trends in Cognitive Sciences*, *15*, 358–364.

Durstewitz, D., Seamans, J. K., & Sejnowski, T. J. (2000). Neurocomputational models of working memory. *Nature Neuroscience*, *3*, 1184–1191.

Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *Science*, *329*, 1541–1543.

Galvin, S. J., Podd, J. V., Drga, V., & Whitmore, J. (2003). Type 2 tasks in the theory of signal detectability: Discrimination between correct and incorrect decisions. *Psychonomic Bulletin & Review*, *10*, 843–876.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York, NY: Wiley.

Kanai, R., Walsh, V., & Tseng, C.-H. (2010). Subjective discriminability of invisibility: A framework for distinguishing perceptual and attentional failures of awareness. *Consciousness and Cognition*, *19*, 1045–1057.

Kouider, S., de Gardelle, V., Sackur, J., & Dupoux, E. (2010). How rich is consciousness? The partial awareness hypothesis. *Trends in Cognitive Sciences*, *14*, 301–307.

Lamme, V. A. F. (2006). Towards a true neural stance on consciousness. *Trends in Cognitive Sciences*, *10*, 494–501.

Lamme, V. A. F. (2010). How neuroscience will change our view on consciousness. *Cognitive Neuroscience*, *1*, 204–220.

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

Lau, H., & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences*, *15*, 365–373.

Lau, H. C., & Passingham, R. E. (2006). Relative blindsight in normal observers and the neural correlate of visual consciousness. *Proceedings of the National Academy of Sciences, USA*, *103*, 18763–18768.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–280.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide*. Hillsdale, NJ: Erlbaum.

Makovski, T., Sussman, R., & Jiang, Y. V. (2008). Orienting attention in visual working memory reduces interference from memory probes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 369–380.

Maniscalco, B., & Lau, H. (2011). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition*, *21*, 422–430.

Metcalfe, J. E., & Shimamura, A. P. (1994). *Metacognition: Knowing about knowing*. Cambridge, MA: MIT Press.

Neisser, U. (1967). *Cognitive psychology*. New York, NY: Apple-Century-Crofts.

Pinto, Y., Sligte, I. G., Shapiro, K., & Lamme, V. A. F. (2013). Fragile visual short-term memory is an object-based and location-specific store. *Psychonomic Bulletin & Review*, *20*, 732–739.

Rahnev, D., Maniscalco, B., Graves, T., Huang, E., de Lange, F. P., & Lau, H. (2011). Attention induces conservative subjective biases in visual perception. *Nature Neuroscience*, *14*, 1513–1515.

Rahnev, D. A., Bahdo, L., de Lange, F. P., & Lau, H. (2012). Prestimulus hemodynamic activity in dorsal attention network is negatively associated with decision confidence in visual perception. *Journal of Neurophysiology*, *108*, 1529–1536.

Rahnev, D. A., Maniscalco, B., Luber, B., Lau, H., & Lisanby, S. H. (2012). Direct injection of noise to the visual cortex decreases accuracy but increases decision confidence. *Journal of Neurophysiology*, *107*, 1556–1563.

Seth, A. K., Dienes, Z., Cleeremans, A., Overgaard, M., & Pessoa, L. (2008). Measuring consciousness: Relating behavioural and neurophysiological approaches. *Trends in Cognitive Sciences*, *12*, 314–321.

Sligte, I. G., Scholte, H. S., & Lamme, V. A. F. (2008). Are there multiple visual short-term memory stores? *PLoS ONE*, *3*(2), e1699. Retrieved from http://www.plosone.org/article/info:doi%2F10.1371%2Fjournal.pone.0001699

Sligte, I. G., Scholte, H. S., & Lamme, V. A. F. (2009). V4 activity predicts the strength of visual short-term memory representations. *The Journal of Neuroscience*, *29*, 7432–7438.

Sligte, I. G., Vandenbroucke, A. R. E., Scholte, H. S., & Lamme, V. A. F. (2010). Detailed sensory memory, sloppy working memory. *Frontiers in Psychology*, *1*, 175. Retrieved from http://www.frontiersin.org/Journal/10.3389/fpsyg.2010.00175/abstract

Sligte, I. G., Wokke, M. E., Tesselaar, J. P., Scholte, H. S., & Lamme, V. A. F. (2011). Magnetic stimulation of the dorsolateral prefrontal cortex dissociates fragile visual short-term memory from visual working memory. *Neuropsychologia*, *49*, 1578–1588.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, *74*, 1–29.

Vandenbroucke, A. R. E., Sligte, I. G., Fahrenfort, J. J., Ambroziak, K. B., & Lamme, V. A. F. (2012). Non-attended representations are perceptual rather than unconscious in nature. *PLoS ONE*, *7*(11), e50042. Retrieved from http://www.plosone.org/article/info:doi/10.1371/journal.pone.0050042

Vandenbroucke, A. R. E., Sligte, I. G., & Lamme, V. A. F. (2011). Manipulations of attention dissociate fragile visual short-term memory from visual working memory. *Neuropsychologia*, *49*, 1559–1568.

Wilimzig, C., Tsuchiya, N., Fahle, M., Einhäuser, W., & Koch, C. (2008). Spatial attention increases performance but not subjective confidence in a discrimination task. *Journal of Vision*, *8*(5), Article 7. Retrieved from http://www.journalofvision.org/content/8/5/7