Cerebral Cortex, 2017; 1–12

doi: 10.1093/cercor/bhx193
Original Article

ORIGINAL ARTICLE

Neural Evidence for Non-conscious Working Memory

Fredrik Bergström[1,2,3] and Johan Eriksson[1,2]

[1]Umeå center for Functional Brain Imaging (UFBI), Umeå University, Sweden, [2]Department of Integrative Medical Biology, Physiology Section, Umeå University, Sweden and [3]Faculty of Psychology and Educational Sciences, University of Coimbra, Portugal

Address correspondence to Johan Eriksson, Department of Integrative Medical Biology, Physiology Section, Umeå University, 901 87 Umeå, Sweden. Email: johan.eriksson@umu.se

Abstract

Recent studies have found that non-consciously perceived information can be retained for several seconds, a feat that has been attributed to non-conscious working memory processes. However, these studies have mainly relied on subjective measures of visual experience, and the neural processes responsible for non-conscious short-term retention remains unclear. Here we used continuous flash suppression to render stimuli non-conscious in a delayed match-to-sample task together with fMRI to investigate the neural correlates of non-conscious short-term (5–15 s) retention. The participants' behavioral performance was at chance level when they reported no visual experience of the sample stimulus. Critically, multivariate pattern analyses of BOLD signal during the delay phase could classify presence versus absence of sample stimuli based on signal patterns in frontal cortex, and its spatial position based on signal patterns in occipital cortex. In addition, univariate analyses revealed increased BOLD signal change in prefrontal regions during memory recognition. Thus, our findings demonstrate short-term maintenance of information presented non-consciously, defined by chance performance behaviorally. This non-consciously retained information seems to rely on persistent neural activity in frontal and occipital cortex, and may engage further cognitive control processes during memory recognition.

Key words: consciousness, continuous flash suppression, fMRI, unconscious, working memory

Introduction

Intuitively, we seem to have rich conscious experiences of our external and internal environment as we navigate our way through the world. However, it is commonly assumed that our conscious experiences reflect but a small fraction of neural processes that occur mainly non-consciously. It was previously assumed that non-conscious processing was simple and automatic, while conscious processing was flexible and strategic (Koch and Crick 2001; Kouider and Dehaene 2007). In recent years, however, there has been a shift in our understanding of non-conscious processing. We now know that non-conscious processing can occur at higher perceptual levels (for reviews see, Rees et al. 2002; Kouider and Dehaene 2007; Koch et al. 2016a), and can influence cognitive control functions in the frontal cortex (Lau and Passingham 2007; van Gaal et al. 2010).

A similar paradigm shift has begun regarding memory. It was previously believed that non-consciously perceived information quickly fades, and is undetectable after 500 ms (Greenwald et al. 1996; Draine and Greenwald 1998; Dehaene and Changeux 2011; Mattler 2005). However, studies have shown that non-conscious repetition priming can have effects lasting 15–20 min (Bar and Biederman 1998, 1999) and even up to 47 min (Gaillard et al. 2007). It was furthermore assumed that non-conscious (non-procedural) memory only existed in the form of priming since working memory, as well as long-term retention involving the hippocampus, were strongly associated with conscious experience (Graf and Schacter 1985; Squire et al. 1992; Baddeley and Andrade 2000; Dehaene and Naccache 2001; Tulving 2002; Baars and Franklin 2003; Baddeley 2003; Baars 2005; Dehaene and Changeux 2011; Squire and Dede 2015). However, there are now several studies showing hippocampus-based retention of non-consciously encoded information (Henke et al. 2003; Degonda et al. 2005; Reber et al. 2012; Chong et al. 2014; Duss et al. 2014).

Lately, the dominating view that working memory only pertains to conscious information has also been challenged (Soto and Silvanto 2014). Working memory is the temporary retention of information for prospective use (Baddeley and Hitch 1974; Baddeley 1983; Fuster 1995, 2015), and has typically been associated with persistent neural activity in the prefrontal cortex related to the task at hand, and in posterior regions related to the memorandum (for reviews, see Fuster 2009; Sreenivasan et al. 2014; Eriksson et al. 2015). Several studies report that non-conscious information can be retained for durations up to 15 s, even with distractors occurring between the sample and memory probe (Soto et al. 2011; Bergström and Eriksson 2014, 2015). This short-term retention of non-consciously perceived information has been associated with sustained BOLD signal change in the prefrontal cortex during retention (Bergström and Eriksson 2014), and activity in prefrontal cortex has been causally linked to task performance using transcranial direct current stimulation (Dutta et al. 2014). Moreover, Pan et al. (2013) demonstrated that non-conscious retention depended on whether or not the information was needed for prospective action, a key feature of working memory. However, there has been some critique against these findings (Samaha 2015; Stein et al. 2016), in that subjective ratings of conscious experience have been used in most previous studies, which can be biased towards under-reporting. That is, participants may have reported “no visual experience” when actually having a “vague visual experience” on some of the trials. Objective measures of performance provide more conservative evidence for stimuli being non-conscious.

To further verify the phenomenon of working memory of non-conscious stimuli and to investigate how the brain accomplishes such retention, we here used continuous flash suppression (CFS) to render stimuli non-conscious while participants performed a delayed match-to-sample task during fMRI scanning. Based on previous research on both conscious and non-conscious working memory, we hypothesized that the neural mechanisms of non-conscious working memory would be weaker but similar to that of conscious working memory, and therefore expected to find sustained BOLD signal change in the prefrontal cortex and possibly posterior sample-specific regions during the delay phase (Bergström and Eriksson 2014; Sreenivasan et al. 2014). Multivariate pattern analyses were used as a more sensitive analysis technique to complement the standard univariate approach, and to provide further information on the type of representations maintained during task performance (Lewis-Peacock and Postle 2008; Lewis-Peacock et al. 2012).

Materials and Methods

Participants

Thirty-four participants were recruited from the Umeå University campus. All participants had normal or corrected to normal vision, right eye-dominance, were right-handed, gave written informed consent, and were paid for participation. The experiment consisted of 2 sessions: 1 pre-fMRI and 1 fMRI session. The pre-fMRI session was used to screen for participants for whom CFS worked well (i.e., reporting target stimuli as unseen on >80% of the suppressed trials). Four participants were excluded prior to the fMRI session for experiencing the target stimulus on too many trials. Two participants were excluded from analyses of behavioral data from the pre-fMRI session for systematically pressing “no match” (instead of guessing) when not experiencing the target stimulus, but were instructed to change their behavior before the fMRI session. Four participants were excluded from the fMRI session: two for failing to follow task instructions, one for excessive head motion, and one for being a statistical outlier in both sessions (pre-fMRI d'= –0.59, > 2 SD; fMRI d' = –0.66, > 2 SD below group mean) when not experiencing target stimuli. Thus, 25 participants (M = 25 years, 17 females) were included in the pre-fMRI session, and 26 participants (M = 25 years, 17 females) in the fMRI session.

Stimuli and Procedure

The pre-fMRI session consisted of 360 delayed match-to-sample trials divided into 3 presentation conditions: 120 conscious, 180 non-conscious, and 60 trials with no target, hereafter referred to as baseline trials (Fig. 1). The fMRI session consisted of 192 delayed match-to-sample trials (44 conscious, 108 non-conscious, and 40 baseline trials). Each trial was drawn randomly from the 3 conditions and began with an intertrial-interval (ITI; 3–7 s for the pre-fMRI session, 3–9 s for the fMRI session) before the sample presentation. The sample consisted of a gray silhouette of a tool (pre-fMRI: 1.7° × 1.7°, fMRI: 1.5° × 1.5°, Gaussian blur: one pixel radius) that was presented in one quadrant of a computer screen. The tool (selected from a set of 6 tools) and quadrant was semirandomly selected, such that the tool and quadrant was not directly repeated from one trial to the next. A mirror stereoscope was used to isolate visual input from one side of the screen to the participants corresponding eye. The screen was placed such that all visual input could be presented within 6° horizontally and 9.6° vertically for the pre-fMRI, and 5.4° horizontally and 8.7° vertically for the fMRI session. The sample was presented for 3 s, either to both eyes simultaneously (consciously experienced), or only to the non-dominant (left) eye while colored squares of random composition (mondrians; pre-fMRI: 4.2° × 4.2°, fMRI: 3.8° × 3.8°) where flashed (10 Hz) to the dominant eye to suppress the sample from conscious experience (Tsuchiya and Koch 2005). During the suppressed presentations the sample was presented at gradually stronger contrast within the initial 400 ms of the 3 s to facilitate suppression. During the baseline trials mondrians were presented to the dominant eye while an empty gray background (pre-fMRI: 4.2° × 4.2°, fMRI: 3.8° × 3.8°) was presented to the non-dominant eye. Critically, the visual experience of baseline and non-conscious trials was the same (experiencing only mondrians).

After a delay phase (5 s during the pre-fMRI session, 5–15 s for the fMRI session), a memory probe was presented until the participant responded (maximum 5 s). The probe could match the sample in terms of object identity and spatial position, only object identity, only spatial position, or neither. The participants were instructed that in this memory recognition task a “match” consisted of the probe being the same object in the same spatial position (i.e., full match). If the probe contained the same object at a different spatial position (identity match), different object at the same spatial position (position match), or different object identity at a different spatial position (non-match), it should be answered with a “no match” response. If participants did not experience a target stimulus (i.e., only experienced mondrians) they were instructed to guess on the first alternative that came to mind/gut feeling (match or no match) when the probe appeared. During the pre-fMRI session there were equal proportions of “match” and “no match” trials. However, during the fMRI session there was a larger proportion of match than no match trials because we intended to focus our analyses on comparisons between hits > baseline and hits > misses. Out of the

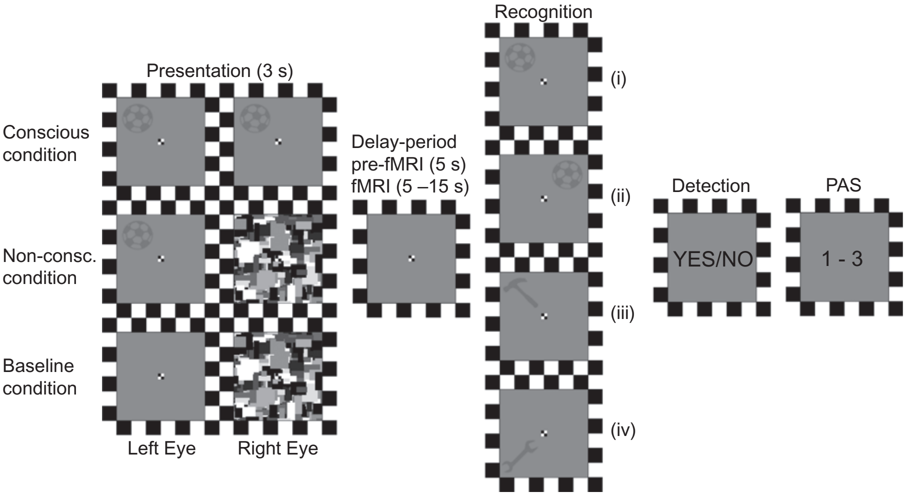

Figure 1. Trial procedures. Depending on the presentation condition, 2 identical target samples (tools), a sample and mondrians (colored, here illustrated in black and white), or an empty background and mondrians, were presented to the left and right eye, respectively. The object identity and spatial position of the sample was then to be retained during a 5 s (pre-fMRI session) or variable 5–15 s (fMRI session) delay phase, until a probe prompted the participants to respond whether or not the probe's identity and position matched the previously presented sample. Next, participants responded whether or not a sample had been present. Finally, the participants gave an estimate of their perceptual experience of the sample. PAS = perceptual awareness scale, (i) probe identity and position matches sample, (ii) probe identity matches sample, (iii) probe position matches sample, and (iv) probe does not match sample.

conscious trials there were 20 full match, 8 identity match, 8 position match, and 8 non-match trials (i.e., 24 "no match" trials). Out of the non-conscious trials there were 78 full match, 10 identity match, 10 position match, and 10 non-match trials.

Next the participants were prompted to make a detection response to determine if a sample stimulus had been presented at all (yes or no). If they had not experienced seeing a sample they were to guess per the same instructions as for the memory-recognition task. Lastly, they estimated their conscious experience of the sample on a three-point perceptual awareness scale (PAS; Sandberg et al. 2010). The participants were instructed and trained to use the PAS scale as follows: 1 = no perceptual experience, 2 = vague perceptual experience, and 3 = clear or almost clear perceptual experience, of the sample stimulus. During the pre-fMRI session there was no time limit when responding, but during the fMRI session responses had an upper time limit of 5 s, after which the experiment automatically continued with the next response prompt or trial.

After participants had received instructions during the pre-fMRI session, they performed a practice run of the experiment with the instructor until their behavior was consistent with the instructions, after which the actual pre-fMRI session began. Participants were debriefed and asked about their behavior in relation to the instructions after both sessions.

fMRI Acquisition

The MRI data were collected with a GE 3 Tesla Discovery MR750 scanner (32-channel receive-only head coil). Each participant underwent one fMRI session with 2 functional runs (1230 volumes each) of scanning using a T2*-weighted gradient echo pulse sequence, echo planar imaging, field of view = 25 cm, matrix size = 96 × 96, slice thickness = 3.4 mm, 37 slices with no interslice skip and an ASSET acceleration factor of 2. The volumes covered the whole cerebrum and most of the cerebellum, the acquisition orientation was oblique axial and aligned with the anterior and posterior commissures, and were scanned in interleaved order with TE = 30 ms, TR = 2 s, flip angle = 78°. Between the 2 functional runs a high-resolution T1-weighted structural image was collected FSPGR with TE = 3.2 ms, TR = 8.2 ms, TI = 450 ms, and flip angle = 12°.

Data Processing and Statistical Analysis

Trials with a response time (RT) of <250 ms or >5 s were excluded prior to statistical analyses (Ratcliff 1993). Only trials in the baseline and non-conscious presentation conditions with PAS = 1, and trials in the conscious condition with PAS = 3 were used in the statistical analyses, and will for simplicity hereby be referred to as baseline, non-conscious, and conscious trials. Signal detection theory (d') was used to calculate performance on the delayed match-to-sample recognition task and the detection task (Macmillan and Creelman 1991). For recognition d' the signal was defined as the object identity and its spatial position. A hit was therefore defined as a (identity and position) match between sample and probe together with a "match" response, and a false alarm (FA) as a non- or partial-match between sample and probe together with a "match" response. For the detection task, a hit was defined as the presence of a sample stimulus together with a "yes" response, and a FA as the absence of a sample stimulus (i.e., baseline trials) together with a "yes" response.

Preprocessing and Univariate Analyses of fMRI Data

The software used for processing and analysis of fMRI data was SPM8 (Welcome Trust Centre for Neuroimaging, London, UK), run in Matlab 7.11 (Mathworks, Inc., Sherborn, MA, USA). Before preprocessing, a manual quality inspection using in-house software was done. Preprocessing was done in the following order; slice-timing correction to the first slice using a Fourier phase-shift interpolation method, head-motion correction with unwarping of B0 distortions, DARTEL normalization (Ashburner 2007) using a 12-parameter affine transformation model to MNI anatomical space, and an 8 mm FWHM Gaussian smoothing. DARTEL normalization and smoothing was applied on the contrast images after intrasubject model estimation. For intrasubject modeling a General Linear Model (GLM) with restricted maximum likelihood estimation was used.

The model consisted of the following regressors of interest: Presentation conditions (conscious and non-conscious) by trial phases (sample presentation, delay, and recognition response) by PAS rating (1, 2, or 3) by signal-detection category (hits, misses, false alarms, and correct rejections), and baseline by trial phases by PAS rating, and lastly the ITI. The model also included the following nuisance regressors: missed responses (because of time limit), head motion (6 parameters) and physiological noise (6 parameters) estimated with temporal variation in white matter and cerebrospinal fluid (Behzadi et al. 2007). All regressors except for head motion and physiological noise were convolved with the "canonical" hemodynamic response function as defined in SPM8. The high-pass filter had a cut-off at 128 s, and the autocorrelation model was global AR(1). Model estimations from each individual were taken to second-level random-effects analyses (one-sample t-tests) to account for interindividual variability. Statistical inferences were made on the whole brain with $P \leq 0.001$, uncorrected for multiple comparisons, $k \geq 20$, unless otherwise specified.

Multivariate Pattern Analyses of fMRI Data

The fMRI data was preprocessed by correcting for slice timing and head motion, as described above for the univariate analysis, prior to being analyzed with the Princeton MVPA Toolbox. For individual feature selection we created binary masks from univariate F contrasts of conscious (hits and correct rejections) compared with baseline trials for the phase of interest (presentation, delay, or recognition response). The feature-selection masks were threshold at $P \leq 0.0001$, uncorrected, $k = 0$ for the whole brain classifications of all phases except recognition ($P \leq 0.001$), which otherwise excluded almost all voxels for some individuals. The feature-selection masks were used to identify voxels relevant for the processing related to the specific phase of interest. We thus assumed that voxels related to conscious processing are relevant for non-conscious processing (Dehaene et al. 2001; Moutoussis and Zeki 2002, 2006; Degonda et al. 2005).

For the presentation- and delay phases we also constructed spatially limited feature-selection masks, defined as the conjunction between the previously described univariate F contrasts ($P \leq 0.001$) and occipital-, temporal-, parietal-, or frontal-lobe masks based on the WFU Pickatlas (Maldjian et al. 2003). Classification during the ITI was used as a "sanity check", and was performed on BOLD signal from the scan prior to sample-presentation onset.

To identify relevant time points with consideration of the slow development of the hemodynamic response we used the regressors from the univariate GLM. Different cut-off values were used in order to only include fMRI data from when the regressors were at their respective peaks: >0.4 for presentation, and >0.14 for recognition. For the delay phase, we did not use the GLM regressors but instead only included trials with a delay duration >10 s and also shifted the onset times forward four scans (8 s). We also cut the period short (relative the prolonged hemodynamic response) so that it always ended at the scan prior to recognition-probe onset, to ensure that no BOLD signal from the memory probe could influence classification performance of the delay. Thus, classification analyses of the delay phase were only conducted on BOLD signal at the end of each delay phase (10–15 s) up until probe onset.

We used 2 different classification analyses to investigate non-conscious representations. In the first analysis we trained a classifier to discriminate between trials where a sample stimulus was presented but suppressed from conscious experience with CFS (PAS = 1), and trials where sample stimuli were absent (baseline trials, PAS = 1). This analysis therefore identifies signal change related to presence/absence of a memory sample, regardless of specific sample features (e.g., tool, position). In the second analysis we trained a classifier to discriminate between suppressed target stimuli presented in the left and right visual field, i.e., a specific sample feature. We chose spatial position rather than object identity because position could easily be grouped into 2 categories (left and right visual field) for more power, and should be relatively easily detected if present because of the retinotopic organization of the visual system.

The included voxel values were passed through a high-pass filter (128 s cut-off), replaced by z-score normalized values, and averaged over time. The analyses used a leave-k-out cross-validation procedure where k is the number of categories to be classified (i.e., 2; either sample presence vs. absence or left vs. right visual field). When there were more trials of one category we randomly excluded trials from that category until there was an equal amount of trials in each category. Because the MVPA needed an equal amount of trials in each category and there were fewer baseline than non-conscious trials, we could only use a subset of all non-conscious trials when classifying presence versus absence of suppressed sample stimuli. We therefore opted to use non-conscious hits and correct rejections versus baseline trials, in case correct trials carried any additional information despite behavioral performance being at chance. When classifying spatial position we used all non-conscious trials except for the random exclusion done to make sure there was an equal amount of trials in both categories.

To preclude that finger-related BOLD signal confounded the classification performance during analyses of the recognition phase, we controlled for which finger/button was used to make the response (index finger for "match", and middle finger for "non-match" responses). For example, if classifying hits and correct rejections versus baseline trials, an equal number of hits and baseline trials with "match" responses, and the same amount of correct rejections and baseline trials with "non-match" responses, were randomly selected.

Following Polyn et al. (2005), we used a backpropagation-based neural network algorithm to train and test the BOLD signal patterns in the data, an OnOff-value was calculated as a measure of classification performance and statistical significance was tested using a non-parametric permutation test. Each participant's OnOff value was computed by correlating the 2 classifier estimates (how well the current test pattern matches each category's characteristic/trained pattern) with the actual conditions (answer key) for all test iterations. From the resulting two-by-two correlation matrix an OnOff value was derived by subtracting the average of the off-diagonal elements from the average of the on-diagonal elements. An overall OnOff value was estimated by averaging across participants' OnOff values. For the non-parametric permutation (group level) test we scrambled the individual OnOff matrices. The actual overall OnOff value was then compared with a null distribution of 10 000 scrambled OnOff values to generate a one-tailed P-value.

Results

In the following results, all trials with PAS > 1 were removed to ensure no visual experience of the target stimulus in non-conscious (3.9% for the pre-fMRI session and <1% for the fMRI session) and baseline conditions (pre-fMRI: 3.3%; fMRI: 0), and all trials with PAS < 3 were removed from the conscious condition (pre-fMRI: 2.5%, fMRI: 2.3%).

Behavioral Performance

One-tailed *t*-tests were used to test whether memory performance (d') was above chance ($d' > 0$) for the recognition and detection tasks. For conscious trials, memory performance was above chance during both sessions for the recognition task [pre-fMRI: $t(24) = 86$, $P < 0.001$, $M = 4.56$, SE = 0.06; fMRI: $t(25) = 89$, $P < 0.001$, $M = 3.85$, SE = 0.04] and the detection task [pre-fMRI: $t(23) = 33$, $P < 0.001$, $M = 4.32$, SE = 0.13; fMRI: $t(25) = 113$, $P < 0.001$, $M = 4.05$, SE = 0.04]. For non-conscious trials, recognition [$t(24) = 3.44$, $P = 0.001$, $M = 0.16$, SE = 0.05] and detection [$t(23) = 3.49$, $P = 0.001$, $M = 0.15$, SE = 0.04] was significantly better than chance during the pre-fMRI session. Importantly, neither recognition [t(25) = 0.41, P = 0.69, $M = 0.02$, SE = 0.04] nor detection [t(25) = 0.31, P = 0.76, $M = 0.02$, SE = 0.05] was better than chance during the fMRI session. Thus, in the fMRI trials the sample to be remembered was non-conscious according to objective criteria.

Similar results were evident for response times. One-tailed paired *t*-tests of conscious recognition response times (ms) during the pre-fMRI session showed that hits [$t(24) = -7.40$, $P < 0.001$, $M = 1236$, SE = 41] and correct rejections [$t(24) = -4.84$, $P < 0.001$, $M = 1505$, SE = 47] were faster than baseline response times ($M = 2027$, SE = 113), and hits were faster than correct rejections [$t(24) = -8.82$, $P < 0.001$]. This was also true for the fMRI session (hits [1287 ± 41] < correct rejections [1517 ± 64] < baseline [2074 ± 105], all $P < 0.001$). For non-conscious recognition during the pre-fMRI session, hit (2015 ± 119), correct rejection (2030 ± 108), and baseline (2027 ± 113) response times did not differ from each other (all $P > 0.49$). However, response times were significantly slower [$t(24) = 2.85$, $P = 0.005$] when participants' guesses were incorrect ($M = 2107$, SE = 108) compared with correct ($M = 2028$, SE = 107), and when compared with baseline trials [$t(24) = 2.00$, $P = 0.029$]. The slower response time was also significant when comparing misses to hits [$t(24) = 2.25$, $P = 0.017$] and false alarms to correct rejections [$t(24) = 1.82$, $P = 0.04$], the latter indicating that the slowing of incorrect responses cannot entirely be accounted for by repetition priming (no stimulus repetition between sample and probe).

For non-conscious recognition during the fMRI session there was no difference in response times between hits, correct rejections, and baseline trials (all $P > 0.27$), nor was there a slower response time for incorrect guesses (all $P > 0.26$). Previous research has demonstrated different neurocognitive processes during working-memory retrieval depending on whether the memory probe matched or did not match the sample (Bledowski et al. 2012; Rahm et al. 2014; Schon et al. 2016). We therefore compared the response times of non-conscious match and non-match trials. Paired *t*-tests revealed that match response-times [$t(25) = 0.24$, $P = 0.81$, $M = 2081$, SE = 90] and non-match response-times [$t(25) = 1.42$, $P = 0.17$, $M = 2119$, SE = 105] were no different from baseline, or each other [$t(25) = -0.95$, $P = 0.35$]. Thus, there were no response-time differences among non-conscious trial types during the fMRI session.

fMRI Results

We examined the BOLD signal change related to conscious processes by contrasting conscious (hits and correct rejections) to the baseline trials separately at the 3 trial phases: sample presentation, delay, and recognition response. When contrasting conscious > baseline sample presentations we found widespread bilateral BOLD signal change in the occipital pole extending along the lingual and fusiform gyri, inferior occipital and temporal gyri, posterior middle temporal gyri, inferior and superior parietal lobules, and the middle frontal gyrus. For the delay phase there was sustained BOLD signal change bilaterally in the inferior and middle occipital gyri. At memory recognition there was widespread bilateral BOLD signal change in inferior, middle, and superior occipital gyri, inferior and middle temporal gyri, fusiform gyri, inferior and superior parietal lobules, posterior middle frontal gyri, and left middle frontal gyrus. Highly similar patterns were evident when comparing conscious > non-conscious presentation, delay, and recognition phases.

The comparison of conscious sample-probe non-match > match probes was associated with extensive BOLD signal change lateralized in the left posterior inferior and middle temporal gyri, inferior and superior parietal lobules, precentral gyrus extending to the inferior and middle frontal gyri, posterior superior frontal gyrus, and the right putamen/pallidum (Fig. 2A), while comparing match > non-match probes was associated with a higher BOLD signal change in the left middle occipital gyrus.

Since non-conscious recognition and detection performance was at chance, we initially treated hits, misses, false alarms, and correct rejections together when comparing non-conscious to baseline trials for the different trial phases. There was no significant BOLD signal change related to non-conscious > baseline sample presentations, nor during the delay phase. However, during the non-conscious recognition phase BOLD signal increased in the right anterior insula and right inferior frontal gyrus compared with baseline (Table 1). The BOLD signal change in the right inferior frontal gyrus correlated positively with recognition d' [$r(25) = 0.56$, $P = 0.003$, two-tailed], while the correlation between d' and BOLD signal in the anterior insula was at trend [$r(25) = 0.39$, $P = 0.051$, two-tailed].

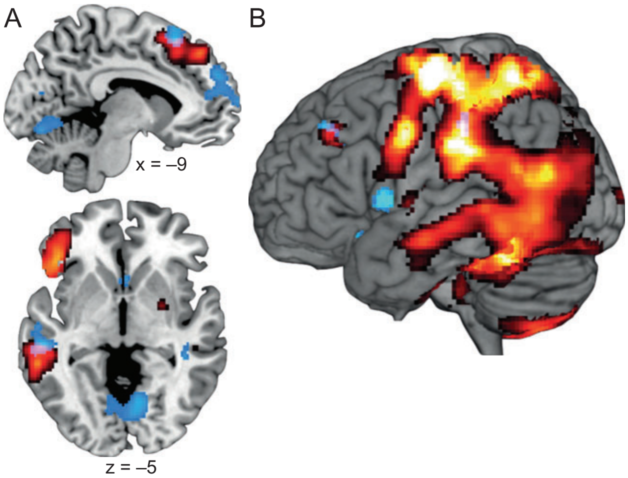

Figure 2. BOLD signal change during memory probe recognition. (*A*) BOLD signal change during conscious (hot colors) and non-conscious (cool colors) trials for memory probe recognition when the probe did not match the sample compared with sample-probe match (non-match > match). Overlap (purple) is evident in medial frontal (upper) and middle temporal (lower) regions. (*B*) BOLD signal change during conscious (hot) and non-conscious (cool) trials for non-matching probes compared with baseline trials. Overlap (purple) is evident in middle frontal and supramarginal gyrus.

Table 1 BOLD signal change during non-conscious memory recognition

Region	Left/Right	Peak *t*-value	XYZ	Cluster size
Non-conscious probe > baseline probe				
Inferior frontal gyrus	R	4.36	[44 14 12]	31
Anterior insula	R	4.09	[32 24 6]	32
Probe match > probe non-match				
Cerebellum	R	4.44	[46 −50 −50]	31
Probe non-match > probe match				
Superior frontal gyrus/medial frontal gyrus **(i)**	L	5.36	[−10 22 54]	189
Superior frontal gyrus	L	4.64	[−8 54 30]	325
Insula	L	4.65	[−40 −6 10]	74
Precentral gyrus	R	4.74	[28 −14 62]	257
	L	4.53	[−32 −14 64]	229
Postcentral gyrus	L	4.39	[−50 −4 40]	85
Superior parietal lobule	L	3.79	[−38 −44 64]	29
Inferior parietal lobule	L	4.90	[−60 −26 38]	458
Middle temporal gyrus **(ii)**	L	5.24	[−54 −28 −4]	240
	L	5.16	[−48 −10 −42]	248
	L	3.84	[−56 −6 −22]	36
	R	4.63	[42 −38 −2]	78
Temporal pole	L	4.11	[−44 20 −18]	44
Calcarine sulcus	L/R	4.04	[12 −82 16]	252
Lingual gyrus	L/R	4.96	[12 −70 −2]	690
Probe non-match > baseline probe				
Middle frontal gyrus **(iii)**	L	3.73	[−36 38 40]	22
Inferior frontal gyrus	L	4.40	[−56 8 4]	45
Medial frontal gyrus	L	4.15	[−10 2 68]	30
Supramarginal gyrus **(iv)**	L	4.03	[−60 −28 40]	54
Cuneus	R	3.83	[8 −84 18]	53
Correct > baseline probe				
Middle frontal gyrus **(v)**	L	3.83	[−38 36 40]	20

Note: (i) and (ii) partly overlapped with conscious probe non-match > match, (iii) and (iv) partly overlapped with conscious probe non-match > baseline (see also Fig. 2), and (v) partly overlapped with (iii) and conscious probe non-match > baseline.

Furthermore, when comparing non-match > match trials there was significant BOLD signal change bilaterally in the lingual gyrus, calcarine sulcus, middle temporal gyrus, and the pre- and postcentral gyrus, as well as left-lateralized signal change in the temporal pole, superior and inferior parietal lobule, insula, superior frontal gyrus, and in the medial frontal gyrus, partly overlapping with the corresponding results for conscious non-match > match (Fig. 2A, Table 1). For non-conscious sample-probe match > non-match, there was a significant BOLD signal change in the right cerebellum (Table 1).

The results related to non-match trials are important because they may demonstrate memory effects that cannot be ascribed to simple repetition between sample and probe. However, the results from the non-match > match comparison may be driven by reduced BOLD signal for match trials (e.g., repetition suppression) rather than BOLD signal change specifically related to non-match trials. To verify that there were significant signal change related specifically to non-match trials we compared non-match > baseline trials, which revealed increased BOLD signal in the left inferior and middle frontal gyrus and supramarginal gyrus, partly overlapping with corresponding results based on conscious trials (Fig. 2B, Table 1). This activity pattern did not overlap with the results from the non-match > match comparison, suggesting that the results from the latter comparison was at least in part driven by signal change related to match trials.

Although behavioral performance during non-conscious recognition was at chance, the significant correlation between *d′* and BOLD signal change in the left inferior frontal gyrus indicates that subtle behavioral effects may be present nonetheless. We therefore also compared non-conscious hits > misses and correct > incorrect trials, but without significant differences for any of the trial phases. Comparing correct > baseline trials during the recognition phase revealed a cluster in the left middle frontal gyrus that overlapped with clusters from conscious and non-conscious non-match > baseline recognition (Table 1). Moreover, the number of hits correlated with BOLD signal (hits > baseline) bilaterally in the superior occipital gyri and superior parietal lobules, the left middle occipital gyrus and cerebellum, the right cuneus, supramarginal gyrus, and postcentral gyrus.

Multivariate Pattern Analyses

The recognition-related results for non-conscious trials demonstrate that some of the non-consciously presented information indeed was retained despite behavioral performance being at chance. It is therefore somewhat surprising that there was no significant signal change at least related to the non-conscious sample presentation. Multivariate pattern analysis (MVPA) is more sensitive than univariate analyses, and previous studies have found significant classification performance without significant univariate BOLD signal change during working-memory tasks (Riggall and Postle 2012; Emrich et al. 2013). We therefore used MVPA to further investigate BOLD signal change during the different trial periods. Specifically, we trained the classifier to differentiate between non-conscious (i.e., suppressed presentations with target stimuli reported as non-conscious; PAS = 1) and baseline trials (i.e., suppressed presentations without target stimuli; PAS = 1) separately for each trial phase (sample presentation, delay, recognition).

Classification accuracy was significantly better than chance (which corresponds to an OnOff value of zero) for the sample-presentation period ($P < 0.0001$, $M = 0.26$, SE = 0.05) using a whole-brain feature mask that was based on conscious versus baseline trials (Fig. 3). To determine if the significant whole-brain analysis was driven by BOLD signal in specific brain regions we combined the whole-brain feature mask with cerebral lobes as defined in the WFU Pickatlas (i.e., the conjunction between the whole-brain feature mask and the different lobes). Classification was successful in each lobe independently (frontal, $P < 0.0001$, $M = 0.32$, SE = 0.06; parietal, $P = 0.009$, $M = 0.15$, SE = 0.06; temporal, $P = 0.003$, $M = 0.19$, SE = 0.06; occipital, $P = 0.036$, $M = 0.09$, SE = 0.05), demonstrating that information regarding sample presence was widely distributed in the brain. To address whether there was any sample-specific information present in the BOLD signal pattern we tried to classify spatial position of the sample object (left vs. right visual field) in regions that are associated with spatial processing, by combining the whole-brain feature mask with cerebral lobes as defined in the WFU Pickatlas. Classification of spatial position was significant in the parietal ($P = 0.01$, $M = 0.13$, SE = 0.05) and occipital ($P = 0.045$, $M = 0.10$, SE = 0.06), but not frontal ($P = 0.25$, $M = 0.04$, SE = 0.05) cortex (Fig. 3).

To investigate non-conscious maintenance processes we used a delay-based feature mask (created from conscious vs. baseline trials) to classify non-conscious sample presence versus absence during the delay phase. Classification performance was not significant ($P = 0.11$, $M = 0.08$, SE = 0.06) when using the whole-brain feature mask. However, subsequent analyses showed that classification was successful in frontal ($P = 0.008$, $M = 0.20$, SE = 0.07), but not parietal ($P = 0.45$, $M = 0.01$, SE = 0.09), temporal ($P = 0.07$, $M = 0.08$, SE = 0.06) or occipital ($P = 0.25$, $M = 0.04$, SE = 0.06) cortex (Fig. 3). The delay phase for the MVPA:s was defined so as to minimize the risk of being affected by residual "spill-over" signal from the sample-presentation phase. To verify that this approach was successful we used the delay-based frontal mask (i.e., the same voxels that was successfully classified during the delay) to classify suppressed sample presence versus absence during the presentation phase. Critically, classification was at chance ($P = 0.31$, $M = 0.03$, SE = 0.07), which excludes the possibility that classification performance for the delay phase was due to residual BOLD signal from the sample presentation.

To address whether sample-specific information was retained during the delay, we tried to decode spatial position. Classification was successful in the occipital ($P = 0.004$, $M = 0.16$, SE = 0.06), but failed in parietal ($P = 0.64$, $M = -0.02$, SE = 0.07), and frontal ($P = 0.15$, $M = 0.08$, SE = 0.07) cortex (Fig. 3). To control for residual BOLD signal we used the delay-based occipital mask to decode spatial position during the sample presentation. Classification performance was at trend ($P = 0.07$, $M = 0.08$, SE = 0.05), which suggests that the relatively high classification performance for the delay phase cannot be wholly attributed to residual BOLD signal from the sample presentation.

Using the recognition-based feature mask (created from conscious vs. baseline trials) classification performance was better than chance for non-conscious sample presence versus absence during recognition ($P < 0.0001$, $M = 0.42$, SE = 0.07), which corroborate our univariate findings and demonstrate that the non-conscious information is retained until recognition despite chance-level performance. Importantly, classification performance was also better than chance when using only non-match trials ($P < 0.0001$, $M = 0.50$, SE = 0.07; Fig. 3).

Control Analyses

It is conceivable that participants accidentally pressed PAS = 1 despite consciously experiencing the target stimulus on some trials, and that such mislabeling could explain the effects attributed to non-conscious processing. We therefore did control analyses to address the potential effect of mislabeling trials in the univariate and multivariate analyses. We assumed that the amount of mislabeling during suppressed trials could be approximated by the number of mislabeled trials in the non-suppressed condition (i.e., trials where the sample stimulus

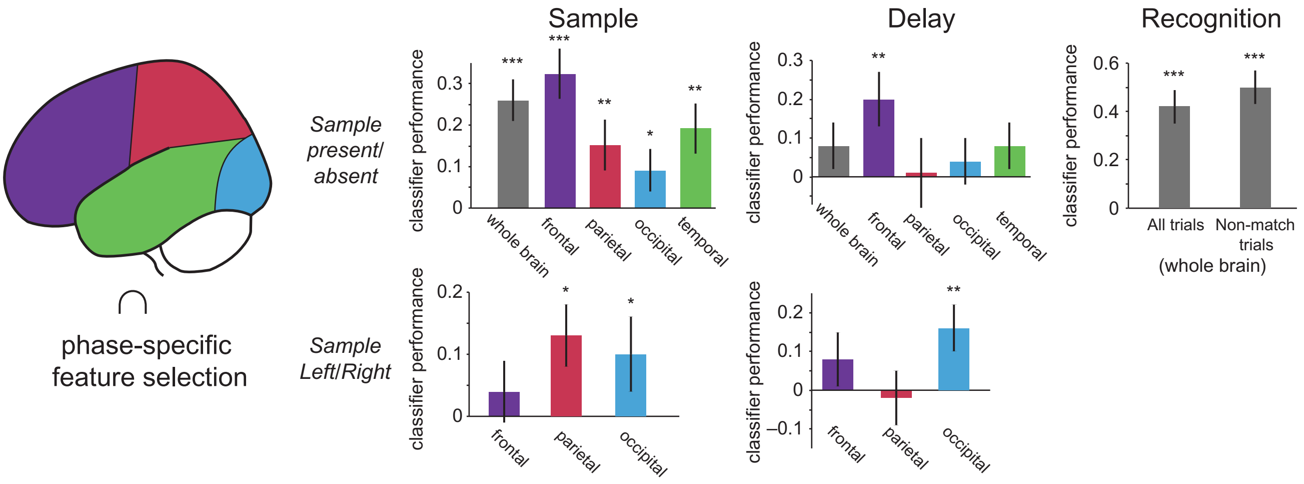

Figure 3. MVPA results. For the MVPA, features (i.e., voxels included in the analyses) were selected based on BOLD signal change during the corresponding trial phase for conscious trials, either across the whole brain (gray bars) or in conjunction with specific cortical lobes (color coded as illustrated in the left panel). Sample presence/absence could be decoded using whole-brain features and each lobe independently during the sample presentation, the whole-brain features during the recognition phase, but only in the frontal lobe during the delay phase. The spatial position of the sample (left/right) could be decoded based on BOLD signal patterns in parietal and occipital cortex during the sample presentation phase, and only in occipital cortex during the delay phase. * = $P < 0.05$, ** = $P < 0.01$, *** = $P < 0.001$; error bars represent one SE; classifier performance = OnOFF values, where zero represents chance performance.

was clearly visible but participants pressed "PAS = 1"; 2.3%, see above). To control for mislabeling effects in the fMRI data we divided the baseline trials in 2 bins. We then contaminated one bin of baseline trials with the corresponding number of potentially mislabeled trials (rounding up, corresponding to one trial, or 5%), using conscious correct-rejection trials (because correct rejections, i.e., non-match trials, had the most pronounced BOLD signal change) for each participant. The univariate comparison of contaminated > pure baseline trials did not reveal any significant BOLD signal change at the previously set threshold, and lowering the threshold to $P \leq 0.01$ uncorrected, $k = 0$, revealed only a few smaller clusters, mostly in white matter, and none that overlapped with previous results. To control for effects of mislabeled trials in the MVPAs we trained the algorithms to differentiate between the contaminated and pure baseline bins. Classification performance was at chance during the sample presentation ($P = 0.37$, $M = 0.03$, SE = 0.08), delay (frontal cortex, $P = 0.20$, $M = 0.06$, SE = 0.07; occipital cortex, $P = 0.68$, $M = -0.04$, SE = 0.08), and recognition phase ($P = 0.46$, $M = 0.01$, SE = 0.09).

Discussion

Consistent with recent research, the current fMRI results demonstrate working memory effects of information presented non-consciously. Here, the effects were most pronounced during the memory recognition (test) phase, but using MVPA it was also possible to decode sample presence and spatial position from BOLD signal patterns during the sample presentation and the delay phase. Critically, behavioral performance during scanning was at chance level, providing strong support for the non-conscious nature of the sample presentation.

Sustained Activity During the Delay Phase

Sustained activity during short memory delays is often considered a characteristic feature of working memory, as it represents a likely neural mechanism for short-term retention of task-relevant information. During the delay phase the MVPA successfully classified presence versus absence of the non-consciously presented sample based on BOLD signal patterns in the frontal cortex, and its spatial position (left vs. right) in the occipital cortex. We cannot infer the exact representational nature of the frontal BOLD signal in relation to retention, except that it is likely not related to spatial information. These findings are consistent with the working-memory literature, which commonly link the prefrontal cortex to task-specific information, while item-specific information (i.e., the memorandum) is suggested to be retained in posterior regions (Fuster 2009; Sreenivasan et al. 2014; Eriksson et al. 2015). It therefore seems like the neural substrates of non-conscious short-term retention is similar to that of conscious working memory. However, given the univariate results, we can infer that non-conscious working memory is much weaker than its conscious counterpart. Surprisingly, it was only possible to decode spatial position and not presence versus absence in the occipital cortex during the delay. Possibly, the BOLD signal during the delay phase was heterogeneous relative present/absent categorization when a sample had been presented, because the signal would toggle between representing left and right samples within the "sample present" category (i.e., different signals within the same category), to a degree that was not apparent during the sample presentation itself. It is also noteworthy that it was not possible to decode spatial position based on BOLD signal in the parietal cortex during the delay phase, even though classification was significant during the sample-presentation phase. Speculatively, only lower-level visuospatial information was actively maintained during the delay.

Considering how weak the BOLD signal was during the delay phase of non-conscious trials (only detectable using MVPA), it is unclear how "sustained" the corresponding neural activity in fact was. Specifically, it may reflect intermittent rather than persistent neural activity (Lundqvist et al. 2015), and/or metabolically demanding synaptic events that may not be reflected in increased neural spiking (Goense and Logothetis 2008), but may still reflect short-term retention of mnemonic information (Shafi et al. 2007; Mongillo et al. 2008). Moreover, we were here able to demonstrate only crude evidence that the sustained information was specifically related to the sample (left/right visual field). These results extend previous findings regarding the type of information that is maintained during the delay phase following non-conscious presentation of a memory sample (King et al. 2016), where successful decoding of target presence/absence was achieved based on MEG signals during the delay phase, but only limited evidence for more specific information was evident. Based on behavioral measures we have previously demonstrated that the conjunction of spatial position and object identity can be maintained during a delayed match-to-sample task almost identical to the current task (Bergström and Eriksson 2015). We here failed to fully replicate these findings and speculate that the extent of encoding during continuous flash suppression may have differed across participants (previous vs. current experiment) and across experimental setups (pre-fMRI vs. fMRI session), such that stronger suppression may lead to a weaker sample representation. Experimentally, the fMRI session was different from the pre-fMRI session and previous behavioral experiments (Bergström and Eriksson 2015) in that we used a different stereoscope due to MR compatibility issues, a longer delay phase to isolate delay-related BOLD signal (5 vs. 5–15 s), and longer experiment time with less rest. All of these factors could have contributed to the decrease in behavioral performance. However, given that we have shown that suppressed information can be retained up to 15 s previously (Bergström and Eriksson 2014, 2015), we speculate that increased drowsiness and reduced task focus during scanning may be a driving factor, as this may specifically affect an active retention mechanism such as working memory. Further research is needed to more clearly determine the kind of information that can be maintained in memory after a non-conscious stimulus presentation, and if such retention is vulnerable to attentional distraction.

Memory Recognition

During the memory recognition phase, there was significant BOLD signal change in the right anterior insula and the right inferior frontal cortex when comparing non-conscious to baseline trials, and the right inferior frontal BOLD signal correlated positively with accuracy. Number of hits correlated with BOLD signal change bilaterally in many posterior regions, notably in the occipital cortex, which previously has been associated with non-conscious recognition performance during null d' in a higher-order implicit sequence learning paradigm (Rosenthal et al. 2016). In line with recent research that specifically address neurocognitive processes during working-memory retrieval (Bledowski et al. 2006, 2012; Nee and Jonides 2008; Rahm et al. 2014), there was also significant activity differences when comparing memory probes that did versus did not match the sample, both for conscious and non-conscious trials. Critically, there was significant BOLD signal change also when comparing only non-matching probes with baseline probes both for the univariate analysis and the MVPA – results that cannot be explained in terms of simple repetition suppression from sample to probe, but rather indicates more complex mnemonic processing that is potentially similar to working-memory retrieval operations during more traditional (i.e., conscious) working-memory tasks. Specifically, previous research has demonstrated increased BOLD signal in left lateral prefrontal cortex during the test phase when demands on sample-probe comparison processes are higher, for example when the memory probe does not match the item currently in the focus of attention (Nee and Jonides 2008; Bledowski et al. 2012; Rahm et al. 2014). Here, BOLD signal change in left middle frontal and supramarginal gyrus increased during non-match trials both when the sample was conscious and non-conscious (Fig. 2*B*), and the left middle frontal signal change was associated with correct guesses.

Several frontal and parietal regions have previously been associated with different cognitive control processes relevant for memory probe recognition, including attentional deployment (Nee and Jonides 2008), sample-probe similarity assessment (Bledowski et al. 2012), and decision making (Rahm et al. 2014). Based on the current activity pattern during the memory-recognition phase and the correlation between BOLD signal in the right inferior frontal gyrus and accuracy, we speculate that cognitive control processes are also engaged during non-conscious memory recognition. In line with this proposal, several previous studies have reported activation of cognitive control processes related to non-conscious stimuli (Lau and Passingham 2007; van Gaal et al. 2010; Charles et al. 2013; Reuss et al. 2015). The specific cognitive roles played by the currently activated regions remain unclear, not least because behavioral performance was at chance level.

Consciousness and Working Memory

Historically, working memory and conscious experience have been tightly linked (Soto and Silvanto 2014). Indeed, early working memory models equated its content with conscious experience (James 1890; Atkinson and Shiffrin 1971), but not all contemporary models do. Although some models posit that all working-memory states are conscious (McElree 2006), others postulate additional non-conscious states (Fuster 1995; Cowan 2008), and some remain silent about the relation between working memory and conscious experience (Oberauer 2002). It has recently been demonstrated that unattended information during working memory tasks can fail to be decoded with MVPA during delay phases, but can nevertheless be used for solving the task, and can be “revived” by re-directing attention (Lewis-Peacock et al. 2012; LaRocque et al. 2013) to the relevant representation or by the application of TMS (Rose et al. 2016). Such findings are in line with state-based models of working memory, where memory representations are in different “states of access” depending on attentional deployment (Fuster 1995; Oberauer 2002; McElree 2006; Cowan 2008; Jonides et al. 2008), but does not address whether the different states are conscious or not. Our current findings suggest that sample representations can be in a heightened state of access, reflected here as significant decoding of BOLD signal patterns during the delay phase, even though they have never been “inside” the focus of attention or consciously experienced. These findings demonstrate that non-conscious representations indeed can be maintained by neural activity over and above activity-silent synaptic changes, and are therefore inconsistent with the “activity-silent” model of non-conscious working memory proposed by Trübutschek et al. (2017) as well as working memory models that only posit conscious states of access (McElree 2006). Working memory is a non-unitary construct whereof short-term information maintenance in the service of ongoing tasks is one key component (Eriksson et al. 2015). The current findings demonstrate that this can be achieved also for non-consciously presented stimuli, and several recent findings demonstrate that other component processes of working memory can also take place non-consciously (see Soto and Silvanto 2014, for overview).

Neuroimaging studies have found overlapping neural correlates of working memory and conscious perception in the prefrontal and parietal cortex (Rees et al. 2002; Naghavi and Nyberg 2005; but see no-report paradigms: Frässle et al. 2014; Tsuchiya et al. 2015; Koch et al. 2016a, 2016b). Based on such findings of overlap, prominent models of conscious experience have asserted that prefrontal and parietal activity, and by extension working memory, plays an important role in conscious experience. According to the Global Neuronal Workspace (GNW) model, non-conscious information is processed locally in specialized modules and is relatively short-lived (<500 ms). For information to become consciously experienced and maintained in working memory, it needs to be globally broadcast via the frontoparietal network, and is thereby available to many brain regions (Dehaene and Naccache 2001; Dehaene and Changeux 2011). However, our findings seem to contradict some of the GNW model’s assumptions. Firstly, non-conscious information can be retained for several seconds by persistent frontal and occipital activity. Secondly, conscious experience of the memorandum does not seem to be necessary for working memory. Thirdly, the presence of non-consciously presented samples could be decoded based on BOLD signal patterns in each cerebral lobe independently, which demonstrates that non-conscious perception can lead to a very global activity pattern, and extends previous findings showing (limited) frontoparietal involvement in non-conscious cognition (Kranczioch et al. 2005; Diaz and McCarthy 2007; Lau and Passingham 2007; van Gaal et al. 2010; Bergström and Eriksson 2014). Recently, King et al. (2016) also demonstrated widespread patterns of brain activity associated with non-conscious working memory using MEG and decoding analyses. Interestingly, the activity pattern that characterized non-conscious working memory differed from the pattern that characterized conscious working memory. Thus, while we here find crude similarities between conscious and non-conscious working memory, associated processes may differ at finer scales. Overall, the current and previous findings of non-conscious processes associated with widespread brain patterns are consistent with the commonly held notion that most of our cerebral processing are parallel and non-conscious, and that we only consciously experience a small fraction of it all. For example, we do not consciously experience all of the processing underlying our visual perception, language production, and sensory–motor reflexes and skills (Zeki 1994, 2001; Koch and Crick 2001; Koch 2004).

Critique of Non-Conscious Working Memory

Recent critique against findings related to non-conscious working memory have pointed out that subjective measures of awareness, which has been used in most previous research on this topic, might be biased towards under-reporting (Samaha 2015; Stein et al. 2016). Results may therefore be driven by information that was in fact conscious, despite subjective reports indicating no conscious experience. Objective measures are more conservative but increase the risk of false negative findings, and the 2 approaches may be seen as complementary (Seth et al. 2008). Stein et al. (2016) argued that even if a subjective measure indeed was to be bias-free, results could still be explained by participants having a non-conscious perception, but this is then transformed into a conscious "guess" that can be maintained in (conscious) working memory. By contrast, several aspects of the current findings support the notion of non-conscious working memory. Firstly, the participants' recognition and detection performance was at chance during the fMRI session, meaning that the sample presentation was non-conscious according to objective criteria. This finding is consistent with results reported by Pan et al. (2013), where memory performance related to non-consciously presented faces was at chance level while indirect measures (time to breaking suppression) was significantly altered by the sample presentation. Secondly, we performed control analyses with regard to accidental mislabeling of trials. These control analyses (univariate and multivariate fMRI) showed that accidental mislabeling at frequencies similar to mislabeling of conscious trials could not by itself drive the effects seen for non-conscious trials. Thirdly, the participants were instructed to wait until the probe appeared before making their guesses, and when debriefed about any particular strategies during the non-conscious trials, they reported none. If, as Stein et al. (2016) suggest, participants guessed the object's identity and position and held that conscious guess in working memory, conscious and non-conscious response times should not differ. However, response times for conscious trials were significantly faster than non-conscious trials, which in turn did not differ from baseline trials. This suggests that participants had not already made their guesses before the probe appeared. Taken together, the current findings provide strong support for the notion of non-conscious working memory.

Conclusions

In conclusion, we found neural evidence for maintenance of non-consciously presented information during several seconds and engagement of brain regions associated with cognitive control during memory recognition. The maintenance of sample-unspecific information in the frontal cortex and sample-specific information in the occipital cortex is consistent with current conceptions of how information is maintained in working memory. These findings imply that working-memory models need to accommodate a representational state where information can be maintained without ever being inside the focus of attention. Furthermore, the findings contradict some of the assumptions of the global neuronal workspace model, namely that non-conscious processing cannot be global, maintained in working memory, or engage cognitive control processes.

Funding

This work was supported by grants from the Swedish Research Council, the medical faculty and the collecting foundation for medical research at Umeå University (to J.E.). The research leading to these results has received funding from the European Union Seventh Framework Program (FP7/2007–13) under grant agreement no. 604102 (Human Brain Project), to Lars Nyberg.

Notes

We thank Lars Nyberg for input on the experimental design, Göran Westling for building the stereoscopes, and Anders Lundqvist and Micael Andersson for statistical advice. *Conflict of Interest*: None declared.

References

Ashburner J. 2007. A fast diffeomorphic image registration algorithm. Neuroimage. 38:95–113.

Atkinson RC, Shiffrin RM. 1971. The control of short-term memory. Sci Am. 225:82–90.

Baars B. 2005. Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. Brain. 150:45–53.

Baars B, Franklin S. 2003. How conscious experience and working memory interact. Trends Cogn Sci. 7:166–172.

Baddeley A. 2003. Working memory: looking back and looking forward. Nat Rev Neurosci. 4:829–839.

Baddeley AD, Andrade J. 2000. Working memory and the vividness of imagery. J Exp Psychol Gen. 129:126–145.

Baddeley A, Hitch GJ. 1974. Working memory. In: Bower GA, editor. Recent advances in learning and motivation. New York: Academic Press. p. 47–89.

Baddeley AD. 1983. Working memory. Philos Trans R Soc B. 302: 311–324.

Bar M, Biederman I. 1998. Subliminal visual priming. Psychol Sci. 9:464–468.

Bar M, Biederman I. 1999. Localizing the cortical region mediating visual awareness of object identity. Proc Natl Acad Sci U S A. 96:1790–1793.

Behzadi Y, Restom K, Liau J, Liu TT. 2007. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. Neuroimage. 37:90–101.

Bergström F, Eriksson J. 2014. Maintenance of non-consciously presented information engages the prefrontal cortex. Front Hum Neurosci. 8:1–10.

Bergström F, Eriksson J. 2015. The conjunction of non-consciously perceived object identity and spatial position can be retained during a visual short-term memory task. Front Psychol. 6:1–9.

Bledowski C, Cohen Kadosh K, Wibral M, Rahm B, Bittner RA, Hoechstetter K, Scherg M, Maurer K, Goebel R, Linden D. 2006. Mental chronometry of working memory retrieval: a combined functional magnetic resonance imaging and event-related potentials approach. J Neurosci. 26:821–829.

Bledowski C, Kaiser J, Wibral M, Yildiz-Erzberger K, Rahm B. 2012. Separable neural bases for subprocesses of recognition in working memory. Cereb Cortex. 22:1950–1958.

Charles L, van Opstal F, Marti S, Dehaene S. 2013. Distinct brain mechanisms for conscious versus subliminal error detection. Neuroimage. 73:80–94.

Chong TT-J, Husain M, Rosenthal CR. 2014. Recognizing the unconscious. Curr Biol. 24:R1033–R1035.

Cowan N. 2008. What are the differences between long-term, short-term, and working memory? In: Progress in brain research. Amsterdam: Elsevier.

Degonda N, Mondadori CR a, Bosshardt S, Schmidt CF, Boesiger P, Nitsch RM, Hock C, Henke K. 2005. Implicit associative learning engages the hippocampus and interacts with explicit associative learning. Neuron. 46:505–520.

Dehaene S, Changeux J-P. 2011. Experimental and theoretical approaches to conscious processing. Neuron. 70:200–227.

Dehaene S, Naccache L. 2001. Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. Cognition. 79:1–37.

Dehaene S, Naccache L, Cohen L, Bihan DL, Mangin JF, Poline JB, Rivière D. 2001. Cerebral mechanisms of word masking and unconscious repetition priming. Nat Neurosci. 4: 752–758.

Diaz MT, McCarthy G. 2007. Unconscious word processing engages a distributed network of brain regions. J Cogn Neurosci. 19:1768–1775.

Draine SC, Greenwald AG. 1998. Replicable unconscious semantic priming. J Exp Psychol Gen. 127:286–303.

Duss SB, Reber TP, Hänggi J, Schwab S, Wiest R, Müri RM, Brugger P, Gutbrod K, Henke K. 2014. Unconscious relational encoding depends on hippocampus. Brain. 137:3355–3370.

Dutta A, Shah K, Silvanto J, Soto D. 2014. Neural basis of non-conscious visual working memory. Neuroimage. 91:336–343.

Emrich SM, Riggall AC, Larocque JJ, Postle BR. 2013. Distributed patterns of activity in sensory cortex reflect the precision of multiple items maintained in visual short-term memory. J Neurosci. 33:6516–6523.

Eriksson J, Vogel EK, Lansner A, Bergström F, Nyberg L. 2015. Neurocognitive architecture of working memory. Neuron. 88:33–46.

Frässle S, Sommer J, Jansen A, Naber M, Einhauser W. 2014. Binocular rivalry: frontal activity relates to introspection and action but not to perception. J Neurosci. 34:1738–1747.

Fuster JM. 1995. Memory in the Cerebral cortex – an empirical approach to neural networks in the human and nonhuman primate. Cambridge, MA: MIT Press.

Fuster JM. 2009. Cortex and memory: emergence of a new paradigm. J Cogn Neurosci. 21:2047–2072.

Fuster JM. 2015. The prefrontal cortex. 5th ed. London: Academic Press.

Gaillard R, Cohen L, Adam C, Clemenceau S, Hasboun D, Baulac M, Willer J-C, Dehaene S, Naccache L. 2007. Subliminal words durably affect neuronal activity. Neuroreport. 18: 1527–1531.

Goense JBM, Logothetis NK. 2008. Neurophysiology of the BOLD fMRI signal in awake monkeys. Curr Biol. 18:631–640.

Graf P, Schacter DL. 1985. Implicit and explicit memory for new associations in normal and amnesic subjects. J Exp Psychol Learn Mem Cogn. 11:501–518.

Greenwald AG, Draine SC, Abrams RL. 1996. Three cognitive markers of unconscious semantic activation. Science. 273: 1699–1702. (80-).

Henke K, Treyer V, Nagy ET, Kneifel S, Dürsteler M, Nitsch RM, Buck A. 2003. Active hippocampus during nonconscious memories. Conscious Cogn. 12:31–48.

James W. 1890. The principles of psychology. New York: Henry Holt and Company.

Jonides J, Lewis RL, Nee DE, Lustig CA, Berman MG, Moore KS. 2008. The mind and brain of short-term memory. Annu Rev Psychol. 59:193–224.

King J-R, Pescetelli N, Dehaene S. 2016. Brain mechanisms underlying the brief maintenance of seen and unseen sensory information. Neuron. 92:1122–1134.

Koch C. 2004. The quest for consciousness: a neurobiological approach. Englewood, Colorado: Roberts & Company Publishers.

Koch C, Crick F. 2001. The zombie within. Nature. 411:893.

Koch C, Massimini M, Boly M, Tononi G. 2016a. Neural correlates of consciousness: progress and problems. Nat Rev Neurosci. 17:307–321.

Koch C, Massimini M, Boly M, Tononi G. 2016b. The neural correlates of consciousness: progress and problems. Nat Rev Neurosci. 17:307–321.

Kouider S, Dehaene S. 2007. Levels of processing during non-conscious perception: a critical review of visual masking. Philos Trans R Soc Lond B Biol Sci. 362:857–875.

Kranczioch C, Debener S, Schwarzbach J, Goebel R, Engel AK. 2005. Neural correlates of conscious perception in the attentional blink. Neuroimage. 24:704–714.

LaRocque JJ, Lewis-Peacock J a, Drysdale AT, Oberauer K, Postle BR. 2013. Decoding attended information in short-term memory: an EEG study. J Cogn Neurosci. 25:127–142.

Lau HC, Passingham RE. 2007. Unconscious activation of the cognitive control system in the human prefrontal cortex. J Neurosci. 27:5805–5811.

Lewis-Peacock J a, Drysdale AT, Oberauer K, Postle BR. 2012. Neural evidence for a distinction between short-term memory and the focus of attention. J Cogn Neurosci. 24:61–79.

Lewis-Peacock J a, Postle BR. 2008. Temporary activation of long-term memory supports working memory. J Neurosci. 28:8765–8771.

Lundqvist M, Rose J, Herman P, Brincat SL, Buschman TJ, Miller EK. 2015. Gamma and beta bursts underlie working memory. Neuron. 90:152–164.

Macmillan NA, Creelman CD. 1991. Detection theory: a user's guide. New York: Cambridge University Press.

Maldjian JA, Laurienti PJ, Kraft RA, Burdette JH. 2003. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. Neuroimage. 19: 1233–1239.

Mattler U. 2005. Inhibition and decay of motor and nonmotor priming. Percept Psychophys. 67:285–300.

McElree B. 2006. Accessing recent events. Psychol Learn Motiv – Adv Res Theory. 46:155–200.

Mongillo G, Barak O, Tsodyks M. 2008. Synaptic theory of working memory. Science. 319:1543–1546.

Moutoussis K, Zeki S. 2002. The relationship between cortical activation and perception investigated with invisible stimuli. Proc Natl Acad Sci U S A. 99:9527–9532.

Moutoussis K, Zeki S. 2006. Seeing invisible motion: a human FMRI study. Curr Biol. 16:574–579.

Naghavi HR, Nyberg L. 2005. Common fronto-parietal activity in attention, memory, and consciousness: shared demands on integration? Conscious Cogn. 14:390–425.

Nee DE, Jonides J. 2008. Neural correlates of access to short-term memory. Proc Natl Acad Sci U S A. 105:14228–14233.

Oberauer K. 2002. Access to information in working memory: exploring the focus of attention. J Exp Psychol Learn Mem Cogn. 28:411–421.

Pan Y, Lin B, Zhao Y, Soto D. 2013. Working memory biasing of visual perception without awareness. Atten Percept Psychophys. 76:2051–2061.

Polyn SM, Natu VS, Cohen JD, Norman KA. 2005. Category-specific cortical activity precedes retrieval during memory search. Science. 310:1963–1966.

Rahm B, Kaiser J, Unterrainer JM, Simon J, Bledowski C. 2014. FMRI characterization of visual working memory recognition. Neuroimage. 90:413–422.

Ratcliff R. 1993. Methods for dealing with reaction time outliers. Psychol Bull. 114:510–532.

Reber TP, Luechinger R, Boesiger P, Henke K. 2012. Unconscious relational inference recruits the hippocampus. J Neurosci. 32:6138–6148.

Rees G, Kreiman G, Koch C. 2002. Neural correlates of consciousness in humans. Nat Rev Neurosci. 3:261–270.

Reuss H, Kiesel A, Kunde W. 2015. Adjustments of response speed and accuracy to unconscious cues. Cognition. 134:57–62.

Riggall AC, Postle BR. 2012. The relationship between working memory storage and elevated activity as measured with functional magnetic resonance imaging. J Neurosci. 32: 12990–12998.

Rose NS, LaRocque JJ, Riggall AC, Gosseries O, Starrett MJ, Meyering EE, Postle BR. 2016. Reactivation of latent working memories with transcranial magnetic stimulation. Science. 354:1136–1139.

Rosenthal CR, Andrews SK, Antoniades CA, Kennard C, Soto D. 2016. Learning and recognition of a non-conscious sequence of events in human primary visual cortex. Curr Biol. 26: 834–841.

Samaha J. 2015. How best to study the function of consciousness? Front Psychol. 6:1–3.

Sandberg K, Timmermans B, Overgaard M, Cleeremans A. 2010. Measuring consciousness: is one measure better than the other? Conscious Cogn. 19:1069–1078.

Schon K, Newmark RE, Ross RS, Stern CE. 2016. A working memory buffer in parahippocampal regions: evidence from a load effect during the delay period. Cereb Cortex. 26: 1965–1974.

Seth AK, Dienes Z, Cleeremans A, Overgaard M, Pessoa L. 2008. Measuring consciousness: relating behavioural and neurophysiological approaches. Trends Cogn Sci. 12:314–321.

Shafi M, Zhou Y, Quintana J, Chow C, Fuster J, Bodner M. 2007. Variability in neuronal activity in primate cortex during working memory tasks. Neuroscience. 146:1082–1108.

Soto D, Mäntylä T, Silvanto J. 2011. Working memory without consciousness. Curr Biol. 21:R912–R913.

Soto D, Silvanto J. 2014. Reappraising the relationship between working memory and conscious awareness. Trends Cogn Sci. 18:520–525.

Squire LR, Dede AJO. 2015. Conscious and unconscious memory systems. Cold Spring Harb Perspect Biol. 7:a021667.

Squire LR, Kosslyn S, Zola-Morgan S, Haist F, Musen G. 1992. Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. Psychol Rev. 99:195–231.

Sreenivasan KK, Curtis CE, D'Esposito M. 2014. Revisiting the role of persistent neural activity during working memory. Trends Cogn Sci. 1–8.

Stein T, Kaiser D, Hesselmann G. 2016. Can working memory be non-conscious?. Neurosci Conscious. 2016(1):niv011.

Trübutschek D, Marti S, Ojeda A, King J-R, Mi Y, Tsodyks M, Dehaene S. 2017. A theory of working memory without consciousness or sustained activity. Elife. 6:1–46.

Tsuchiya N, Koch C. 2005. Continuous flash suppression reduces negative afterimages. Nat Neurosci. 8:1096–1101.

Tsuchiya N, Wilke M, Frässle S, Lamme VAF. 2015. No-report paradigms: extracting the true neural correlates of consciousness. Trends Cogn Sci. 19:757–770.

Tulving E 2002. Episodic Memory: From Mind to Brain.

van Gaal S, Ridderinkhof KR, Scholte HS, Lamme VAF. 2010. Unconscious activation of the prefrontal no-go network. J Neurosci. 30:4143–4150.

Zeki S 1994. A Vision of the Brain, Optometry and Vision Science.

Zeki S. 2001. Localization and globalization in conscious vision. Annu Rev Neurosci. 24:57–86.