

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25

**A theory of working memory**  
**without consciousness or sustained activity**

Darinka Trübutschek<sup>1, 2, 3, \*</sup>, Sébastien Marti<sup>3</sup>, Andrés Ojeda<sup>4</sup>, Jean-Rémi King<sup>5, 6</sup>, Yuanyuan  
Mi<sup>7</sup>, Misha Tsodyks<sup>8, 9</sup>, and Stanislas Dehaene<sup>3, 10</sup>

- 1 – Ecole des Neurosciences de Paris Ile-de-France, 15 rue de l'Ecole de médecine, 75006 Paris, France
- 2 – Université Pierre et Marie Curie, 4 Place Jussieu, 75005 Paris, France
- 3 – Cognitive Neuroimaging Unit, CEA DSV/I2BM, INSERM, Université Paris-Sud, Université Paris-Saclay, NeuroSpin center, 91191 Gif/Yvette, France
- 4 – Department of Zoology, University of Oxford, Oxford OX1 3PS, United Kingdom
- 5 – Department of Psychology, New York University, 4 Washington Place, New York, USA
- 6 – Frankfurt Institute for Advanced Studies, Frankfurt, Germany
- 7 – Brain Science Center, Institute of Basic Medical Sciences, 100850 Beijing, China
- 8 – Department of Neurobiology, Weizmann Institute of Science, 76100 Rehovot, Israel
- 9 – Department of Neuroscience, Columbia University, 10032 New York, USA
- 10 – Collège de France, 11 Place Marcelin Berthelot, 75005 Paris, France

\*Correspondence: [darinkat87@gmail.com](mailto:darinkat87@gmail.com)

26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40

## Abstract

Working memory and conscious perception are thought to share similar brain mechanisms, yet recent reports of non-conscious working memory challenge this view. Combining visual masking with magnetoencephalography, we investigate the reality of non-conscious working memory and dissect its neural mechanisms. In a spatial delayed-response task, participants reported the location of a subjectively unseen target above chance-level after several seconds. Conscious perception and conscious working memory were characterized by similar signatures: a sustained desynchronization in the alpha/beta band over frontal cortex, and a decodable representation of target location in posterior sensors. During non-conscious working memory, such activity vanished. Our findings contradict models that identify working memory with sustained neural firing, but are compatible with recent proposals of ‘activity-silent’ working memory. We present a theoretical framework and simulations showing how slowly decaying synaptic changes allow cell assemblies to go dormant during the delay, yet be retrieved above chance-level after several seconds.

41

## Introduction

42 Prominent theories of working memory require information to be consciously  
43 maintained (Baars and Franklin, 2003; Baddeley, 2003; Oberauer, 2002). Conversely,  
44 influential models of visual awareness hold information maintenance as a key property of  
45 conscious perception, highlighting synchronous thalamocortical activity (Tononi and Koch,  
46 2008), cortical recurrence (Lamme and Roelfsema, 2000), or the sustained recruitment of  
47 parietal and dorsolateral prefrontal regions (i.e., the same areas as in working memory;  
48 Naghavi and Nyberg, 2005) in a global neuronal workspace (Dehaene and Changeux, 2011;  
49 Dehaene and Naccache, 2001). Experimentally, non-conscious priming only lasts a few  
50 hundred milliseconds (Dupoux et al., 2008; Greenwald et al., 1996) and unseen stimuli  
51 typically fail to induce late and sustained cerebral responses (Dehaene et al., 2014). Conscious  
52 perception, in contrast, exerts a durable influence on behavior, accompanied by sustained  
53 neural activity (King et al., 2014; Salti et al., 2015; Schurger et al., 2015). The hypothesis of  
54 an intimate coupling between conscious perception and working memory is thus grounded in  
55 theory and supported by numerous empirical findings.

56 Recent behavioral and neuroimaging evidence, however, has questioned this  
57 prevailing view by suggesting that working memory may also operate non-consciously.  
58 Unseen stimuli may influence behavior for several seconds (Bergström and Eriksson, 2015;  
59 Soto and Silvanto, 2014). Soto and colleagues (2011), for instance, showed that participants  
60 recalled the orientation of a subjectively unseen Gabor cue above chance-level after a 5s-  
61 delay. Functional magnetic resonance imaging suggests that prefrontal activity may underlie  
62 such non-conscious working memory (Bergström and Eriksson, 2014; Dutta et al., 2014).

63 The verdict for non-conscious working memory is far from definitive, however.  
64 Delayed performance with subjectively unseen stimuli was barely above chance (Soto et al.,  
65 2011) and could have arisen from a small percentage of errors in visibility reports, with

66 subjects miscategorizing a seen target as unseen (miscategorization hypothesis). If this were  
67 the case, then the blindsight trials, on which subjects correctly identified the target while  
68 denying any subjective awareness of the stimulus, should display similar, if not identical,  
69 neural signatures and contents as the seen trials. Alternatively, participants could also have  
70 ventured a guess about the target as soon as it appeared and consciously maintained this early  
71 guess (conscious maintenance hypothesis). Many priming studies have shown that fast  
72 guessing results in above-chance objective performance with subjectively unseen stimuli  
73 (Merikle, 2001). The observed blindsight effect would then reflect a normal form of conscious  
74 working memory (Stein et al., 2016). This alternative hypothesis is hard to eliminate on  
75 purely behavioral grounds; it can only be rejected by tracking the dynamics of working  
76 memory activity, for instance using brain-imaging, and determining whether this activity  
77 occurs immediately after the target even on unseen trials.

78 Here, we set out to address these issues, focusing on four main objectives: First, we  
79 probed the replicability of the long-lasting blindsight effect reported by Soto et al. (2011) as  
80 well as its robustness with respect to interference from distraction and a conscious working  
81 memory load in order to delineate it from other forms of prolonged iconic or sensory memory.  
82 Second, we interrogated the link between conscious perception and conscious working  
83 memory, examining whether the maintenance period in working memory could be likened to  
84 a prolongation of a conscious episode. Third, we tested the reality of non-conscious working  
85 memory by systematically examining the neural correlates of the blindsight effect and using  
86 them to assess the two above alternative hypotheses (the miscategorization and conscious  
87 maintenance hypothesis). Lastly, we propose a neuronal theory to offer a mechanistic account  
88 of conscious and non-conscious working memory.

## 89 **Results**

90 We combined magnetoencephalography (MEG) with a spatial masking paradigm to  
91 assess working memory performance under varying levels of subjective visibility (Figure 1A  
92 and Methods). On 80% of the trials, a target square was flashed in 1 of 20 locations and then  
93 masked. Subjects were asked to localize the target after a variable delay (2.5 – 4.0s) and to  
94 rate its visibility on a scale from 1 (not seen) to 4 (clearly seen). On the remaining 20% of  
95 trials, the target was omitted, allowing us to contrast brain activity between target-present and  
96 -absent trials. A visible distractor square was presented 1.5s into the delay period on half the  
97 trials, challenging participants' resistance to distraction and enabling us to evaluate the  
98 robustness of the blindsight effect behaviorally. In addition to this working memory task,  
99 subjects also completed a perception-only control condition without the delay and target-  
100 localization periods (perception task), so that we could isolate brain activity specific to  
101 conscious perception (without a working memory requirement) and investigate its link with  
102 working memory.

### 103 **Behavioral maintenance and shielding against distraction**

104 We first examined objective performance in the working memory task as a function of  
105 target visibility. Overall, subjects reported the exact target location with high accuracy on  
106 seen trials (collapsed across visibility ratings  $> 1$ :  $M_{correct} = 69.1\%$ ,  $SD_{correct} = 17.4\%$ ; chance  
107  $= 5\%$ ;  $t(16) = 15.2$ ,  $p < .001$ , 95% CI = [55.2%, 73.1%]; Cohen's  $d = 3.7$ ). As subjective  
108 visibility of the target increased from glimpsed (visibility = 2) to clearly seen (visibility = 4),  
109 there was a corresponding monotonic increase in accuracy (Figure 1B;  $ps < .05$  for all pair-  
110 wise comparisons). Crucially, performance remained above chance even on unseen trials  
111 (rating = 1:  $M_{correct} = 22.4\%$ ,  $SD_{correct} = 13.8\%$ ;  $t(16) = 5.2$ ,  $p < .001$ , 95% CI = [10.3%,  
112 24.4%]; Cohen's  $d = 1.3$ ). This blindsight remained substantial after a 4s-delay ( $M_{correct} =$   
113  $21.1\%$ ,  $SD_{correct} = 14.7\%$ ;  $t(16) = 4.5$ ,  $p < .001$ , 95% CI = [8.5%, 23.7%]; Cohen's  $d = 1.0$ ).

114 Spatial distributions of participants' responses were concentrated around the target  
115 (Figure 2A). To correct for small errors in localization, we computed the rate of correct  
116 responding with a tolerance of two positions ( $\pm 36^\circ$ ) surrounding the target location. In  
117 subjects displaying above-chance blindsight (chance = 25%;  $p < .05$  in a  $\chi^2$ -test;  $n = 13$ ), we  
118 estimated the precision of working memory as the standard deviation of the distribution  
119 within this tolerance interval (Methods). Performance was better on seen than on unseen  
120 trials, both in terms of rate of correct responding ( $F(1, 16) = 198.5, p < .001$ ; partial  $\eta^2 =$   
121  $.925$ ) and precision ( $F(1, 12) = 36.7, p < .001$ ; partial  $\eta^2 = .754$ ). There was neither an effect  
122 of the distractor on these measures (all  $ps > .079$ ), nor any significant interactions between  
123 distractor and visibility (all  $ps > .251$ ), indicating that distractor presence did not affect  
124 retention for seen or unseen targets. Restricting the analyses to trials within one position of  
125 the actual target location ( $\pm 18^\circ$ ) or to the subgroup of 13 subjects included in the MEG  
126 analyses did not change these findings qualitatively.

127 While target detection  $d'$  exceeded chance-level ( $M = 1.5, SD = 0.7; t(16) = 8.9, p <$   
128  $.001$ , 95% CI = [1.2, 1.9]; Cohen's  $d = 2.1$ ) and correlated with accuracy and the rate of  
129 correct responding on seen trials (both Pearson  $rs > .762$ , both  $ps < .001$ ), there was no  
130 relationship between our participants' sensitivity to the target and any of our performance  
131 measures on the unseen trials (all Pearson  $rs < .342$ , all  $ps > .179$ ; Figure 2 – Figure  
132 Supplement 1A). Thus, target visibility predicted performance in the objective working  
133 memory task only on seen trials, but not on unseen trials.

134 Overall, these results confirm, with much higher non-conscious performance, the  
135 observations of previous studies (Soto et al., 2011): Non-conscious information may be  
136 maintained for up to 4 seconds and successfully shielded against distraction from a salient  
137 visual stimulus, independently of overall subjective visibility.

138 **Resistance to conscious working memory load and delay duration**

139 To probe the similarity between conscious working memory and the observed long-  
140 lasting blindsight effect, in a second behavioral experiment with 21 subjects, we examined  
141 whether imposing a load on conscious working memory (remembering digits) affected non-  
142 conscious performance. On each trial, 1 (low load) or 5 (high load) digits were simultaneously  
143 shown for 1.5s, followed by a 1s-fixation period and the same sequence of events (target and  
144 mask) as in experiment 1. After a variable delay (0 or 4s), participants had to (1) localize the  
145 target, (2) recall the digits in the correct order, and (3) rate target visibility.

146 Subjects again chose the exact target position with high accuracy on seen trials  
147 ( $M_{correct} = 77.8\%$ ,  $SD_{correct} = 13.9\%$ ) and remained above chance on unseen trials ( $M_{correct} =$   
148  $25.6\%$ ,  $SD_{correct} = 11.8\%$ ; chance = 5%;  $t(18) = 7.6$ ,  $p < .001$ , 95% CI = [14.9%, 26.3%];  
149 Cohen's  $d = 1.7$ ). While, as in experiment 1, cue detection  $d'$  was greater than chance ( $M =$   
150  $1.7$ ,  $SD = 0.8$ ;  $t(20) = 10.2$ ,  $p < .001$ , 95% CI = [1.4, 2.1]; Cohen's  $d = 2.2$ ), no correlations  
151 were observed with objective task performance on the unseen trials (all Pearson  $r$ s  $< .366$ , all  
152  $p$ s  $> .115$ ; seen trials: all Pearson  $r$ s  $> .443$ , all  $p$ s  $< .051$ ; Figure 2 – Figure Supplement 1B).  
153 As expected, participants were better at recalling 1 rather than 5 digits in the correct order ( $M$   
154  $= 93.3\%$  vs.  $89.5\%$ ,  $F(1, 17) = 4.7$ ,  $p = .045$ ), irrespective of target visibility or delay duration  
155 (all  $p$ s  $> .135$ ).

156 Analyzing only the trials with correctly recalled digits, we observed an impact of load  
157 on the precision with which target location was retained ( $F(1, 13) = 7.3$ ,  $p = .018$ ; partial  $\eta^2 =$   
158  $.360$ ). Crucially, load modulated the relationship between precision and visibility (interaction  
159  $F(1, 13) = 8.7$ ,  $p = .011$ ; partial  $\eta^2 = .400$ ), with no effect on seen ( $t(13) = 0.6$ ,  $p = .561$ ) and a  
160 strong reduction of precision on unseen trials ( $t(13) = -3.6$ ,  $p = .004$ ). There was no effect of  
161 working memory load on the rate of correct responding (all  $p$ s  $> .229$ ; Figure 2B).

162 Delay duration (0 or 4s) also did not influence the rate of correct responding (all  $p$ s  $>$   
163  $.082$ ; Figure 2C). It did, however, affect overall precision ( $F(1, 15) = 9.3$ ,  $p = .008$ ; partial  $\eta^2$

164 = .383) and the relationship between precision and visibility (interaction  $F(1, 15) = 5.2, p =$   
165  $.037$ ; partial  $\eta^2 = .259$ ). This interaction was driven by higher precision on no-delay than on  
166 4s-delay trials, exclusively when subjects had seen the target ( $t(15) = -5.7, p < .001$ ; unseen  
167 trials:  $t(15) = -0.6, p = .559$ ).

168 Overall, these results highlight the replicability and robustness of the long-lasting  
169 blindsight effect and suggest that it does not just constitute a prolonged version of iconic  
170 memory: Even in the presence of a concurrent conscious working memory load, unseen  
171 stimuli could be maintained, with no detectable decay as a function of delay. However, the  
172 systems involved in the short-term maintenance of conscious and non-conscious stimuli  
173 interacted, because a conscious verbal working memory load diminished the precision with  
174 which non-conscious spatial information was maintained.

### 175 **Similarity of conscious perception and conscious working memory**

176 To tackle our second objective – a detailed examination of the link between conscious  
177 perception and conscious working memory –, we turned to our MEG data and first ensured  
178 that the mechanisms underlying conscious perception were stable across experimental  
179 conditions. The subtraction of the event-related fields (ERFs) evoked by unseen trials from  
180 those evoked by seen trials revealed similar topographies for the perception and working  
181 memory task (Figure 3A): Starting at ~300ms and extending until ~500ms after target onset, a  
182 response emerged over right parieto-temporal magnetometers. This divergence resulted  
183 primarily from a sudden increase in activity on seen trials (“ignition”) in the perception ( $p_{\text{FDR}}$   
184  $< .05$  from 384 – 416ms and from 504 – 516ms) and working memory task ( $p_{\text{FDR}} < .05$  from  
185 328 – 364ms and from 396 – 404ms; Figure 3B). The observed topographies and time courses  
186 fall within the time window of typical neural markers of conscious perception, including the  
187 P3b (e.g., Del Cul et al., 2007; Salti et al., 2015; Sergent et al., 2005). Consciously perceiving  
188 the target stimulus therefore involved comparable neural mechanisms, irrespective of task.



189           We next directly probed the relationship between conscious perception and  
190 information maintenance in conscious working memory. Does the latter reflect a prolonged  
191 conscious episode, or does it involve a distinct set of processes recruited only during the  
192 retention phase? If conscious working memory can indeed be likened to conscious perception,  
193 one might expect the same patterns that index such perception to be sustained throughout the  
194 working memory maintenance period. Linear multivariate pattern classifiers were trained to  
195 predict visibility (seen or unseen) from MEG signals separately for each task. Classification  
196 performance was assessed during an early time period (100 – 300ms), the critical P3b time  
197 window (300 – 600ms), and the first (0.6 – 1.55s) and second part (1.55 – 2.5s) of the delay  
198 period.

199           Decoding of the visibility effect was comparable in the two tasks (Figure 3C and Table  
200 1): Classification performance rose sharply between 100 and 300ms and peaked during the  
201 P3b time window (all  $ps < .007$ , except 100 – 300ms in the working memory task, where  $p =$   
202  $.066$ ). It then decayed slowly from ~1s onwards in both tasks, yet remained above chance  
203 during the 0.6 – 1.55s interval (all  $ps < .001$ ). Similar time courses were also observed when  
204 training in one task and testing for generalization to the other. Though rapidly dropping to  
205 chance-level after ~1s, classifiers trained in the perception task performed above chance  
206 during the first three time windows on working memory trials (and vice versa; all  $ps < .014$ ),  
207 indicating that, early on, both tasks recruited similar brain mechanisms.

208           Temporal generalization analyses (King and Dehaene, 2014) were used to evaluate the  
209 onset and duration of patterns of brain activity. If working memory were just a prolonged  
210 conscious episode, classifiers trained at time points relevant to conscious perception (e.g., the  
211 P3b window) should generalize extensively, potentially spanning the entire delay. Our  
212 findings supported this hypothesis only in part. The temporal generalization matrix for the  
213 working memory task presented as a thick diagonal, suggesting that brain activity was mainly

214 characterized by changing, but long-lasting patterns. Though failing to achieve statistical  
215 significance over the entire 0.6 – 1.55s interval (all  $ps > .101$ ), at a more lenient, uncorrected  
216 threshold, classifiers trained during the P3b time window (300 – 600ms) in the working  
217 memory task remained weakly efficient until ~692ms (AUC = 0.54 +/- 0.02,  $p_{\text{uncorrected}} =$   
218 .023). Similarly, classifiers trained during the same time period in the perception task and  
219 tested in the working memory task persisted up to ~860ms (AUC = 0.53 +/- 0.01,  $p_{\text{uncorrected}} =$   
220 .028). Brain processes deployed for the conscious representation of the target were thus  
221 partially sustained during the working memory delay. The reverse analysis, in which we  
222 trained classifiers during the retention period in the working memory task (0.8 – 2.5s), did not  
223 reveal any generalization to the P3b time window in the perception task ( $p = .101$ ).

224         These results confirm that seeing the target entailed a similar unfolding of neural  
225 events in two task contexts: Conscious perception primarily consisted in a dynamic series of  
226 partially overlapping information-processing stages, each characterized by temporary,  
227 metastable patterns of neural activity. The same neural codes appeared to be recruited at the  
228 beginning of the maintenance period (up to ~1s). As such, these findings corroborate previous  
229 accounts linking conscious perception to an “ignition” of brain activity (Del Cul et al., 2007;  
230 Gaillard et al., 2009; Salti et al., 2015; Sergent et al., 2005) and suggest that, in part, working  
231 memory implies the prolongation of a conscious episode, and, in part, a succession of  
232 additional processing steps.

### 233 **A sustained decrease in alpha/beta power distinguishes conscious working memory**

234         Our focus so far has been on evoked brain activity. However, other reliable neural  
235 signatures of conscious perception have been identified in the frequency domain (Gaillard et  
236 al., 2009; Gross et al., 2007; King et al., 2016; Wyart and Tallon-Baudry, 2009). We thus  
237 turned to time-frequency analyses and first contrasted seen trials with both our target-absent  
238 control condition as well as unseen trials in both tasks (Figure 4A and Figure 4 – Figure

239 Supplement 1A). In order to qualify as a signature of conscious perception, any candidate  
240 characteristic should exist in the perception-only control condition (without any working  
241 memory requirement) and be specific to seen trials. Cluster-based permutation analyses  
242 singled out a desynchronization in the alpha band (8 – 12Hz) as the principal correlate of  
243 conscious perception in the perception task (seen – target-absent:  $p_{\text{clust}} = .004$ ; seen – unseen:  
244  $p_{\text{clust}} = .009$ ), with seen trials displaying a strong decrease in power (relative to baseline)  
245 compared to either the target-absent or the unseen trials. Initially left-lateralized in centro-  
246 temporal sensors, this effect moved to fronto-central channels and extended between ~300  
247 and 1700ms. A similar, albeit later (500 – 1700ms) and more bilateral fronto-central,  
248 desynchronization was also observed in the beta band (13 – 30Hz; seen – target-absent:  $p_{\text{clust}} <$   
249  $.001$ ; seen – unseen:  $p_{\text{clust}} = .01$ ). No differences between the unseen and target-absent trials  
250 were found in the alpha ( $p_{\text{clust}} > .676$ ) or beta band ( $p_{\text{clust}} > .226$ , apart from a short-lived,  
251 weak difference between ~0.9 and 1.3s, where  $p_{\text{clust}} = .020$ ), suggesting that unseen trials  
252 strongly resembled trials without a target.

253         Most importantly, when comparing seen and target-absent/unseen trials in the working  
254 memory task, we again observed a similar, but now temporally sustained, pattern of  
255 alpha/beta band desynchronization (Figure 4B and Figure 4 – Figure Supplement 1B).  
256 Starting at ~300 to 500ms, seen targets evoked a power decrease in central, temporal/parietal,  
257 and frontal regions in the alpha (seen – target-absent:  $p_{\text{clust}} = .003$ ; seen – unseen:  $p_{\text{clust}} = .003$ )  
258 and beta band (seen – target-absent:  $p_{\text{clust}} = .009$ ; seen – unseen:  $p_{\text{clust}} < .001$ ). Crucially, this  
259 desynchronization spanned the entire delay period and was specific to seen trials (Figure 4A),  
260 with no differences in power between the unseen and target-absent trials in either band (alpha:  
261  $p_{\text{clust}} > .729$ ; beta:  $p_{\text{clust}} > .657$ ) and only a couple of interspersed periods of residual  
262 desynchronization persisting in the target-absent control trials. No task- or visibility-related  
263 modulations in power spectra were found in occipital areas, and the desynchronization

264 originated primarily from a parietal network of brain sources (Figures 4A and B). In  
265 conjunction with the afore-mentioned results, these findings imply that alpha/beta  
266 desynchronization is a correlate of conscious perception (Gaillard et al., 2009) and a neural  
267 state common to conscious perception and conscious working memory.

### 268 **A distinct neurophysiological mechanism for non-conscious working memory**

269 Having identified markers of conscious perception and working memory in both  
270 multivariate and time-frequency analyses, we can now test the reality of non-conscious  
271 working memory by confronting it with several alternative hypotheses. The miscategorization  
272 hypothesis suggests that the long-lasting blindsight resulted from a small set of seen trials  
273 erroneously labeled as unseen. Unseen correct trials should thus display similar neural  
274 signatures as seen trials, including a shared discriminative decoding axis and a  
275 desynchronization in the alpha/beta band. An analogous reasoning holds for the conscious  
276 maintenance hypothesis, according to which the observed blindsight effect arises from the  
277 conscious maintenance of an early guess: Conscious processing would occur on unseen trials  
278 and we should thus find a sustained decrease in alpha/beta power similar to the one on seen  
279 trials. Conversely, a clear distinction between brain responses on seen trials and on unseen  
280 (correct) trials would suggest that blindsight resulted from a distinct non-conscious  
281 mechanism of information maintenance.

282 We first probed the alternative hypotheses with the ERF data. Training a decoder to  
283 distinguish seen from unseen trials in the perception task and applying it to the unseen correct  
284 and incorrect trials in the working memory task, we directly assessed the classifier's ability to  
285 generalize from seen to unseen correct trials (accuracy decoder). If, indeed, the latter had  
286 actually been seen, such a decoder should look similar to the above-described generalization  
287 analysis, in which a classifier had been trained on seen/unseen trials in the perception task and  
288 tested on the same labels in the working memory task (visibility decoder). As shown in Figure

289 4 – Figure Supplement 2A, this was not the case. Whereas the temporal generalization matrix  
290 for the visibility decoder presented as a thick diagonal, no discernable pattern emerged for the  
291 accuracy decoder. The time courses of diagonal decoding were also quite dissimilar. For the  
292 visibility decoder (see also above), classification performance first rose above chance at  
293 ~148ms (AUC = 0.54 +/- 0.01,  $p_{\text{FDR}} = .023$ ), peaked at ~640ms (AUC = 0.58 +/- 0.02,  $p_{\text{FDR}} =$   
294 .001), and then decayed rapidly by ~1s (first three time windows, all  $ps < .001$ ). In contrast,  
295 classification for the accuracy decoder was erratic and transient: It first sharply peaked at  
296 ~180ms (AUC = 0.55 +/- 0.01,  $p_{\text{uncorrected}} = .037$ ), dropped to chance-level, and then exceeded  
297 chance between ~372 and 724ms with a peak at 444ms (AUC = 0.57 +/- 0.02,  $p_{\text{uncorrected}} =$   
298 .007). Much unlike any of the previous decoders involving the perception task, long after the  
299 visibility response, it rose a third time between ~1.44 and 1.74s, peaking with similar  
300 magnitude as before at ~1.58s (AUC = 0.57 +/- 0.02,  $p_{\text{uncorrected}} = .010$ ; P3b and last time  
301 window: all  $ps < .023$ ). Although the level of noise evident in the accuracy decoder thus  
302 precludes any definitive conclusion, the visibility and accuracy decoders had little in  
303 common, rendering it unlikely for the unseen correct trials to have simply been mislabeled.

304 We next turned to time-frequency analysis. When averaging over all unseen trials in  
305 the working memory task, there was no indication of a desynchronization remotely  
306 comparable to the one on seen trials (Figure 4A and Figure 4 – Figure Supplement 1C).  
307 Indeed, Bayesian statistics indicated that, on the unseen trials, evidence for the null hypothesis  
308 (i.e., no relative change in alpha/beta power) was at least similar (at the very end of the epoch)  
309 or stronger than evidence for the alternative hypothesis. By contrast, on seen trials, evidence  
310 for the alternative hypothesis was always strongly favored (Figure 4 – Figure Supplement 3).  
311 Even when analyzing the unseen correct trials separately, there was no appreciable trace of  
312 any alpha/beta desynchronization (Figure 4C and Figure 4 – Figure Supplement 3). Only one  
313 short-lived effect, reversed relative to conscious trials, was observed in the alpha band ( $p_{\text{clust}} =$

314 .040) in a set of posterior central sensors, corresponding to primarily occipital sources:  
315 Starting at ~1.5s and extending until ~1.9s, unseen correct trials exhibited a stronger *increase*  
316 in alpha power than their incorrect counterparts. Given the difference in performance on these  
317 two types of unseen trials, such small variations are not surprising and could, perhaps, reflect  
318 a stronger suppression of interference from the distractor on the unseen correct trials. Unseen  
319 correct trials thus appeared to be nearly indistinguishable from the unseen incorrect and  
320 target-absent trials.

321 As multivariate analyses might be more sensitive than univariate ones in detecting  
322 similarities between conditions, we also performed the above decoding analysis separately for  
323 average alpha (8 – 12 Hz) and beta (13 – 30 Hz) power. Overall, these analyses confirmed our  
324 previous findings, albeit more clearly so in the alpha than in the beta band. A visibility  
325 decoder trained on alpha power to distinguish seen from unseen trials in the perception task  
326 and tested in the working memory task again exhibited a thick diagonal, with above-chance  
327 diagonal decoding between ~180ms and 1.18s (first three time windows: all  $ps < .016$ ). There  
328 was no evidence for any generalization to the unseen correct trials (Figure 4 – Figure  
329 Supplement 2B; all time windows:  $ps > .211$ ). Similarly, a visibility decoder trained on  
330 average beta power entirely failed to generalize to the unseen correct trials (Figure 4 – Figure  
331 Supplement 2C; all time windows:  $ps > .191$ ). Considering the weak, although statistically  
332 significant (all four time windows,  $ps \leq .05$ ), initial generalization from the perception to the  
333 working memory task, probably due to the slightly later onset of the beta desynchronization in  
334 the former, this failure is less informative than the one observed in the alpha band and should  
335 be replicated in future investigations.

336 Taken together, we found a clear distinction in the brain responses of seen and unseen  
337 (correct) trials. Converging evidence from our decoding analyses in the ERFs and alpha/beta  
338 band suggests that there was no apparent discriminative axis shared between the seen and the

339 unseen correct trials. Similarly, the desynchronization in alpha/beta power characterizing the  
340 seen targets did not emerge on the unseen (correct) trials. These findings therefore argue  
341 against the miscategorization and conscious maintenance hypotheses and instead suggest that  
342 non-conscious working memory is a genuine phenomenon, distinct from conscious working  
343 memory.

#### 344 **Contents of conscious and non-conscious working memory can be tracked transiently**

345 We next set out to identify the neural mechanisms supporting both conscious and non-  
346 conscious working memory and first determined where and how the specific contents of  
347 working memory were stored. Circular-linear correlations between the amplitude of the ERFs  
348 and target location (across all working memory trials) revealed a strong and focal association  
349 (relative to a permuted null distribution) over posterior channels, starting at ~120ms and  
350 lasting until 904ms (early and P3b time windows: all  $ps < .001$ ; all BFs  $> 109.60$ ; Figure 5A  
351 and Tables 2 and 3). Similarly, distractor position could be tracked between ~194 and 570ms  
352 after its presentation (early and P3b time windows: all  $ps < .009$ ; all BFs  $> 14.47$ ). The  
353 position of our stimuli could thus be faithfully retrieved in visual areas.

354 In a subsequent step, we investigated how target location would be maintained in the  
355 context of conscious and non-conscious working memory (Figure 5B). Target position was  
356 transiently encoded via slowly decaying activity in occipital as well as bilateral temporo-  
357 occipital cortex from ~120 to 800ms on seen trials (early and P3b time windows: all  $ps < .001$   
358 and all BFs  $> 24.07$ , with the exception of the 100 – 300ms period in right temporo-occipital  
359 channels, where  $p = .064$  and BF = 2.31) and in occipital and left temporo-occipital brain  
360 areas from ~180 to 504ms on unseen trials (early time window: all  $ps < .047$ ; all BFs  $> 2.58$ ).  
361 A clear correlation with target location was therefore found for both seen and unseen trials. In  
362 fact, although it was more short-lived on the latter, it was of comparable magnitude as the one  
363 observed on the seen trials during the early time window (occipital/left temporo-occipital

364 channels: all  $ps > .110$  when directly comparing the correlation scores of seen and unseen  
365 trials in a Wilcoxon signed-rank test). In the case of seen trials, both occipital and left  
366 temporo-occipital cortex also maintained the target representation at least throughout the first  
367 part of the delay period (all  $ps < .024$ ; all BFs  $> 3.77$ ), though, intriguingly, this was not  
368 accompanied by continuously sustained activity. Target “decodability” instead waxed and  
369 waned, appearing and disappearing periodically. No such activity was observed for the  
370 maintenance of unseen targets (first and second part of the delay: all  $ps > .446$ ; all BFs  $<$   
371  $.047$ ). This absence of “decodability” during the maintenance period persisted, even when  
372 considering unseen correct and unseen incorrect trials separately (Figure 5C). There was only  
373 a trace of residual decoding of target location on unseen correct trials in left temporo-occipital  
374 areas during the delay period, but this did not reach significance, potentially due to the low  
375 number of trials in this condition. Note that in the perception task, seen targets could be  
376 retrieved similarly to their counterparts in the working memory task between  $\sim 232$  and  
377  $1184$ ms in occipital and bilateral temporo-occipital regions (all  $ps > .068$ , except for the  $100 -$   
378  $300$ ms time window in occipital channels where  $p = .008$ , when directly comparing the  
379 correlation scores of seen targets in both tasks in a Wilcoxon signed-rank test; Figure 5 –  
380 Figure Supplement 1).

381         Given the univariate nature of the circular-linear correlations, one might again wonder  
382 whether a multivariate strategy would be more sensitive in detecting subtle associations  
383 between the MEG data and target location. We therefore used linear support vector  
384 regressions (SVR) to predict target angle from the MEG signal as a function of visibility  
385 (Methods). As can be seen in Figure 5 – Figure Supplement 2, this method resulted in similar,  
386 albeit more noisy, time courses as the ones obtained with the circular-linear correlations: Seen  
387 targets were again encoded and maintained intermittently between  $\sim 268$ ms and  $1.4$ s (P3b time  
388 window and first part of the delay:  $ps < .05$ ). No statistically significant decoding emerged for



389 unseen target location. Due to the fact that subjects responded correctly on approximately half  
390 of all unseen trials (see Table 4 for average trial counts), we attempted to evaluate the  
391 dynamics of the encoding and maintenance of unseen correct and incorrect target locations by  
392 training the regression model on the strongest case, the seen correct trials, and applying it  
393 separately to the unseen correct and incorrect trials. We again observed no evidence for any  
394 generalization at all (Figure 5 – Figure Supplement 3A), though this likely reflects the  
395 sensitivity of the analysis more so than any meaningful effect.

396         Taken together, in line with previous research (Harrison and Tong, 2009; King et al.,  
397 2016), these results suggest that posterior sensory regions may initially encode seen and  
398 unseen memoranda via slowly decaying neural activity. In the case of conscious working  
399 memory, these then seem to be maintained by those same areas through an intermittently  
400 reactivated, neural code (Fuentemilla et al., 2010). In contrast, no such periodically  
401 resurfacing activity appears to accompany non-conscious working memory.

#### 402 **Further evidence against the conscious maintenance hypothesis**

403         The correlation between target location and brain activity affords an additional way to  
404 interrogate the conscious maintenance hypothesis. If subjects quickly guessed the location of  
405 an unseen target and then held it in conscious working memory, in addition to observing a  
406 signature of conscious processing on the unseen trials, we should observe a correlation with  
407 the location of their response long before it occurs. Potentially, remembering the response  
408 might recruit brain systems completely different from the ones representing the target.

409         Circular-linear correlations rendered this prediction unlikely. Associations between  
410 response location and the MEG signal were again primarily confined to posterior channels,  
411 with more frontal areas being recruited preferentially at the time of the response (Figure 6A).  
412 As such, the topographical patterns were highly similar to the ones observed for the  
413 correlation with target location. Importantly, no additional regions were identified on the

414 unseen trials and none of these areas showed any appreciable correlation before the  
415 presentation of the response screen (Figure 6 – Figure Supplement 1). This suggests that,  
416 irrespective of stimulus visibility, common brain networks supported memories for the target  
417 stimulus and the ensuing decision and that, in the case of non-conscious working memory,  
418 these did not come online until the response.

419         The time courses of the circular-linear correlations further solidified this interpretation  
420 (Figure 6B). On seen trials, response position was maintained throughout the majority of the  
421 epoch in occipital and left temporo-occipital brain areas (first three time windows: all  $ps <$   
422  $.020$ ; all BFs  $> 4.16$ ). This was not the case on the unseen trials: No correlation patterns  
423 appeared in any of the posterior channels during the course of the epoch (all time windows:  
424 all  $ps > .064$ ; all BFs  $< 1.32$ ). In contrast, a strong correlation emerged for both seen and  
425 unseen trials during the response period (0 – 800ms with respect to the onset of the letter cue).  
426 Response location could be tracked with similar time courses and magnitude on seen and  
427 unseen trials in occipital, bilateral temporo-occipital, and frontal channels (all  $ps < .024$ ; all  
428 BFs  $> 13.73$ ; when directly comparing the correlation scores of seen and unseen targets in a  
429 Wilcoxon signed-rank test: all  $ps > .216$ , except for left temporo-occipital channels, where  $p =$   
430  $.040$ ). When we further distinguished unseen correct from unseen incorrect trials, the results  
431 remained similar, though much noisier (Figure 6C): There was no clear correlation pattern  
432 before the onset of the response screen on either the unseen correct or the unseen incorrect  
433 trials (all  $ps > .096$ ; all BFs  $< 1.47$ ). Only after the appearance of the letter cues did we  
434 observe a correlation with response location.

435         Multivariate decoding analyses confirmed this picture: Whereas response location for  
436 seen targets could be tracked similarly to actual target location at least throughout the first  
437 part of the delay period (P3b time window and first part of the delay:  $ps < .05$ ; Figure 6 –  
438 Figure Supplement 2), no such pattern was observed on the unseen trials (all  $ps > .153$ ). This

439 absence of decodability persisted on the unseen correct and incorrect trials, even when  
440 training the regression model on the seen correct trials (Figure 5 – Figure Supplement 3B).

441 Overall, these results are incompatible with the hypothesis that the long-lasting  
442 blindsight is only due to the conscious maintenance of an early guess, as, in this case, brain  
443 responses linked to the subjects' response should have been observed shortly after the  
444 presentation of the target stimulus.

#### 445 **Short-term synaptic change as a neurophysiological mechanism for conscious and non-** 446 **conscious working memory**

447 What mechanism might permit above-chance recall without any continuously  
448 sustained brain activity? Recent modelling suggests that sustained neural firing may not be  
449 required to maintain a representation in conscious working memory. Mongillo, Barak, and  
450 Tsodyks (2008) proposed a theoretical framework for working memory, in which information  
451 is stored in calcium-mediated short-term changes in synaptic weights, thus linking the active  
452 cells coding for the memorized item. Once these changes have occurred, the cell assembly  
453 may go dormant during the delay, while the synaptic weights are slowly decaying. At the end  
454 of the delay period, a non-specific read-out signal may then suffice to reactivate the assembly.  
455 Furthermore, reactivation of the assembly may also occur spontaneously during the retention  
456 phase, similar to the rehearsal process postulated by Baddeley (2003), thus refreshing the  
457 weights and permitting the bridging of longer delays. Could this 'activity-silent' mechanism  
458 also constitute a plausible neural mechanism for non-conscious working memory?

459 To test this hypothesis, we simulated our experiments using a one-dimensional  
460 recurrent continuous attractor neural network (CANN) based on Mongillo et al. (2008). The  
461 CANN encoded the angular position of the target and was composed of neurons aligned  
462 according to their preferred stimulus value (Figure 7A). Transient short-term plasticity  
463 between the recurrent connections, with a 4s-decay constant, was implemented as described

464 by Mongillo and colleagues (2008; Figure 7B). Timing of the simulated events was  
465 comparable to the experimental paradigm: A target signal was briefly presented at a random  
466 location, followed by a mask signal to all neurons and a non-specific recall signal after a 3s-  
467 delay.

468         If the activity-silent mechanism constituted a plausible neurophysiological correlate of  
469 conscious and non-conscious working memory, these simulations should capture our principal  
470 findings. A stimulus presented at threshold should entail one of two different maintenance  
471 regimes: a first distinguished by near-perfect recall with spontaneous reactivations of the  
472 memorized representation throughout the retention period (thus resembling the prolonged, yet  
473 fluctuating, “decodability” of seen target locations), and a second characterized by above-  
474 chance objective performance in the almost complete absence of delay activity (thereby  
475 portraying the time course of the circular-linear correlations for the unseen stimuli).

476         In a noiseless model, there indeed existed a critical value of mask amplitude,  $A_{\text{critical}}$ ,  
477 which separated two distinct regimes: Just as was the case for our seen trials, when  $A_{\text{mask}} <$   
478  $A_{\text{critical}}$ , the neural assembly coding for the target spontaneously reactivated during the delay  
479 (Figure 7C). However, when  $A_{\text{mask}} > A_{\text{critical}}$ , the system evolved into a state without  
480 spontaneous activation of target-specific neurons, yet with a reactivation in response to a non-  
481 specific recall signal, mimicking our unseen trials (Figure 7D). When fixing mask amplitude  
482 near  $A_{\text{critical}}$  and adding noise continuously or just to the inputs, the network exhibited both  
483 types of regimes in nearly equal proportions: 50.8% of trials were characterized by an  
484 activity-silent delay interspersed with spontaneous reactivations and 49.2% by an entirely  
485 activity-silent delay period. Reminiscent of our behavioral results, sorting the trials according  
486 to the existence or absence of these reactivations and computing the histograms of recalled  
487 target position relative to true location produced two distributions of objective working  
488 memory performance: one, in which target position was nearly accurately stored (Figure 7E),

489 and one, in which performance remained above chance despite a higher base rate of errors  
490 (Figure 7F). These simulations replicate our experimental findings (in particular Figures 2 and  
491 5) and suggest the activity-silent framework as a likely candidate mechanism for both  
492 conscious and non-conscious working memory.

## 493 **Discussion**

494 Conscious perception and working memory are thought to be intimately related, yet  
495 recent evidence challenged this assumption by proposing the existence of non-conscious  
496 working memory (Soto et al., 2011). The present results may reconcile these views. Both  
497 conscious perception and conscious working memory shared similar signatures, including an  
498 alpha/beta power decrease, the latter spanning the entire delay on working-memory trials.  
499 However, participants remained able to localize a subjectively invisible target after a 4s-delay.  
500 We found no evidence that this long-lasting blindsight could simply be explained by  
501 erroneous visibility reports or by the conscious maintenance of an early guess. It thus likely  
502 reflects genuine non-conscious working memory. Despite the inherent differences in  
503 subjective experience for conscious and non-conscious working memory, a single, activity-  
504 silent mechanism might support both conscious and non-conscious information maintenance.  
505 We now discuss these points in turn.

### 506 **Shared brain signatures underlie conscious perception and conscious working memory**

507 Consistent with introspective reports and research on visual awareness and working  
508 memory (Baddeley, 2003; Dehaene et al., 2014), we observed a close relationship between  
509 conscious perception and maintenance in conscious working memory. In both tasks,  
510 classifiers trained to separate seen and unseen trials resulted in thick diagonals up to ~1s after  
511 target onset, even when generalizing from one task to the other. Such long diagonals have  
512 repeatedly been observed in recent studies and are thought to reflect sequential processing  
513 (King and Dehaene, 2014; Marti et al., 2015; Salti et al., 2015; Stokes et al., 2015; Wolff et

514 al., 2015). Irrespective of context, conscious perception and early parts of conscious  
515 maintenance thus involve a similar series of partially overlapping processing stages.

516 Time-frequency decompositions reinforced and extended this conclusion. Seen trials  
517 in the perception task were distinguished from both a target-absent control condition and  
518 unseen trials by a prominent decrease in alpha/beta power over fronto-central sensors,  
519 corresponding to a distributed network centered on parietal cortex. A similar  
520 desynchronization, sustained throughout the retention period, was also observed for conscious  
521 working memory. Alpha/beta band desynchronizations such as these have previously been  
522 linked with conscious perception (Gaillard et al., 2009; Wyart and Tallon-Baudry, 2009) and  
523 working memory (Lundqvist et al., 2016). Modelling suggests that the memorized item is  
524 encoded by intermittent gamma bursts, which interrupt an ongoing desynchronized beta  
525 default state (Lundqvist et al., 2011). Such a decreased rate of beta bursts, once averaged over  
526 many trials, would have resulted in the apparently sustained power decrease we observed.  
527 Increases in gamma power have also been shown in some studies on conscious perception  
528 (e.g., Gaillard et al., 2009), but we failed to detect it here, perhaps because our targets were  
529 brief, peripheral, and low in intensity.

530 Circular-linear correlations further highlighted the similarity between conscious  
531 perception and working memory. Location information could be tracked for ~1s on  
532 perception-only trials and for at least 1.5 seconds of the working memory retention period.  
533 The mental representation formed during conscious perception was therefore either  
534 maintained or repeatedly replayed during conscious working memory.

### 535 **Long-lasting blindsight effect reflects genuine non-conscious working memory**

536 Even when subjects indicated not having seen the target, they still identified its  
537 position much better than chance up to 4 seconds after its presentation. This long-lasting  
538 blindsight effect was replicated in two independent experiments and exhibited typical

539 properties of working memory, withstanding salient visible distractors and a concurrent  
540 demand on conscious working memory. Those results corroborate previous research showing  
541 that information can be maintained non-consciously (e.g., Bergström and Eriksson, 2014;  
542 Bergström and Eriksson, 2015; Dutta et al., 2014; Soto et al., 2011). However, these prior  
543 findings could have arisen due to errors in visibility reports. If, for example, a participant had  
544 been left with a weak impression of the target (and, consequently, its location), he or she  
545 might not have had adequate internal evidence to refer to this perceptual state as seen, thus  
546 incorrectly applying the label unseen. A small number of such errors would have produced  
547 above-chance responding. Another explanation could have been the conscious maintenance of  
548 an early guess, whereby subjects would have ventured a prediction as to the correct target  
549 position immediately after its presentation and then consciously maintained this hunch.

550         The MEG results provide evidence against these possibilities. First, whereas seen trials  
551 were characterized by a sustained desynchronization in the alpha/beta band in parietal brain  
552 areas, no comparable desynchronization was observed on unseen trials, even when subjects  
553 correctly identified the target location. On the contrary, the only, short-lived, difference  
554 between unseen correct and unseen incorrect trials emerged around the time of the distractor  
555 and was reversed in direction: Unseen correct trials were accompanied by an increase in  
556 power in the alpha band with respect to their incorrect counterpart, an effect that might relate  
557 to a successful attempt to reduce interference from the distractor (Cooper et al., 2003; Jensen  
558 and Mazaheri, 2010). Otherwise, unseen correct and incorrect trials were indistinguishable in  
559 their power spectra and similar to the target-absent control condition. Second, there was no  
560 clear evidence for a shared discriminative decoding axis between the seen and the unseen  
561 correct trials: Generalization was entirely unsuccessful when the classifier was trained on the  
562 time-frequency data, and highly dissimilar from the original visibility decoder when trained  
563 on the ERFs. While it is impossible to draw definitive conclusions just from the current

564 dataset and future research should replicate these results, the majority of our evidence thus  
565 points against an interpretation, in which the unseen correct trials constituted either just a  
566 subset of seen trials, or arose from the conscious maintenance of an early guess. Instead,  
567 inasmuch as the observed desynchronization serves as a faithful indicator of conscious  
568 processing, it argues in favor of a differential state of non-conscious working memory with a  
569 distinct neural signature.

570         Circular-linear correlations as well as multivariate regression models between the  
571 amplitude of the MEG signal and response location support this interpretation. On seen trials,  
572 response position was coded akin to target location: Initially maintained via slowly decaying  
573 neural activity in posterior brain areas, the response code subsequently resurfaced  
574 intermittently in the same as well as more frontal regions. There was no detectable evidence  
575 for such a code on the unseen trials. Only during the very last part of the delay, right before  
576 the response, did response-related neural activity emerge and ramp up to the same level as on  
577 seen trials during the response period. As such, the absence of any prior delay-period activity  
578 does not appear to be an artifact attributable to low statistical power or an increase in noise on  
579 the unseen trials. Instead, in conjunction with the absence of any signature of conscious  
580 processing on these trials, these findings imply that subjects did not consciously maintain an  
581 early guess and rather relied on genuine non-conscious working memory to perform the task.

582         In this context, an interesting avenue for future investigations might be to delineate the  
583 boundary conditions of such non-conscious working memory. Although the short-term  
584 maintenance of information certainly lies at the heart of most theories of working memory  
585 (Eriksson et al., 2015), there exist additional criteria for working memory that were not  
586 investigated in the present study. It is thus an interesting empirical question whether these  
587 other working memory processes may also occur without subjective awareness. Is it, for  
588 example, possible to manipulate information non-consciously? Though speculative, in light of



589 the proposed activity-silent code for non-conscious maintenance (without any spontaneous  
590 reactivations; see below), it seems unlikely. Being an entirely passive process, it is not clear  
591 how stored representations could be transformed without being persistently activated and thus  
592 becoming conscious. Future research is, however, needed to provide a definitive answer.

### 593 **A theoretical framework for ‘activity-silent’ working memory**

594 Target-related activity was not continuously sustained throughout the delay period,  
595 even when the target square had been consciously perceived. It instead fluctuated,  
596 disappearing and reappearing intermittently. This feature was even more pronounced on the  
597 unseen trials, with no evidence for any such retention-related activity beyond ~1s. We  
598 presented a theoretical framework, based on Mongillo et al. (2008) and the concept of  
599 ‘activity-silent’ working memory (Stokes, 2015), that may provide a plausible explanation for  
600 maintenance without sustained neural activity. According to this model, short-term memories  
601 are retained by slowly decaying patterns of synaptic weights. A retrieval cue presented at the  
602 end of the delay may then serve as a non-specific read-out signal capable of reactivating these  
603 dormant representations above chance-level. Support for this model comes from experiments  
604 in which non-specific, task-irrelevant stimuli (Wolff et al., 2017, 2015), neutral post-cues  
605 (Sprague et al., 2016), or transcranial magnetic stimulation (TMS) pulses (Rose et al., 2016)  
606 presented during a delay restore the decodability of representations. Direct physiological  
607 evidence for the postulated short-term changes in synaptic efficacies also exists (Fujisawa et  
608 al., 2008).

609 The present non-conscious condition provides further support for such an activity-  
610 silent mechanism. In this framework, a stimulus that fails to cross the threshold for sustained  
611 activity and subjective visibility may still induce enough activity in high-level cortical circuits  
612 to trigger short-term synaptic changes. Such transient non-conscious propagation of activity  
613 has indeed been simulated in neural networks (Dehaene and Naccache, 2001) and measured

614 experimentally in temporo-occipital, parietal, and even prefrontal cortices (Salti et al., 2015;  
615 van Gaal and Lamme, 2012). In the present work, we indeed observed some residual,  
616 transiently decodable activity over left occipito-temporal sensors on unseen correct trials. The  
617 memory of target location could therefore have arisen from posterior visual maps (Roelfsema,  
618 2015), although future research should test this prediction further. Note that activity-silent  
619 mechanisms need not apply solely to prefrontal cortex as originally proposed by Mongillo et  
620 al. (2008), but constitute a generic mechanism that may be replicated in different areas,  
621 possibly with increasingly longer time constants across the cortical hierarchy (Chaudhuri et  
622 al., 2014). Only some of these areas/spatial maps may be storing the information on unseen  
623 trials.

624         A key feature of Mongillo et al.'s (2008) model and the present simulations is that,  
625 even for above-threshold ('seen') stimuli, delay activity is not continuously sustained.  
626 Occasional bouts of spontaneous reactivation instead refresh the synaptic weights and  
627 maintain the memory for an indefinite time. The time courses of the circular-linear  
628 correlations and of the multivariate decoding we observed on seen trials match this  
629 description: While target location was encoded and maintained in temporo-occipital areas,  
630 target "decodability" was not constantly sustained, but waxed and waned throughout the  
631 delay. Fuentemilla et al. (2010) also observed that, during a delay period, decodable  
632 representations of memorized images recurred at a theta rhythm. More recently, single-trial  
633 analyses of monkey electrophysiological recordings in a working memory task have  
634 confirmed the absence of any continuous activity and instead identified the presence of  
635 discrete gamma bursts, paired with a decrease in beta-burst probability (Lundqvist et al.,  
636 2016). Such periodic refreshing of otherwise activity-silent representations could potentially  
637 serve as the neural correlate of conscious rehearsal, a central feature of working memory  
638 according to Baddeley (2003). It also suggests, however, that even consciously perceived

639 items may not always be “in mind.” Future research might attempt to more directly simulate  
640 activity-silent mechanisms in the context of conscious and non-conscious perception by, for  
641 example, relying on more elaborate models capturing decreases in alpha/beta (Lundqvist et  
642 al., 2011).

643         In conjunction with prior evidence (King et al., 2016; Salti et al., 2015), our findings  
644 therefore indicate that there may be two successive mechanisms for the short-term  
645 maintenance of conscious and non-conscious stimuli: an initial, transient period of ~1s, during  
646 which the representation is encoded by active firing with a slowly decaying amplitude, and an  
647 ensuing activity-silent maintenance via short-term changes in synaptic weights, during which  
648 activity either intermittently resurfaces (conscious case) or vanishes (non-conscious case).  
649 Such activity-silent retention need not necessarily be specific to working memory. Recent  
650 investigations have, for instance, demonstrated the existence of recognition memory for  
651 invisible cues (Chong et al., 2014; Rosenthal et al., 2016). As delay periods ranged in the  
652 order of minutes rather than seconds, persistent neural activity seems to be an unlikely  
653 candidate mechanism of maintenance. Activity-silent codes might have been at play, though  
654 they probably depended on mechanisms with longer time constants than the relatively rapidly  
655 decaying patterns of synaptic weights discussed in the context of the present experiments.  
656 Nevertheless, activity-silent representations may constitute a general mechanism for  
657 maintenance across the whole spectrum of temporal delays (from seconds over minutes/hours  
658 to days/weeks/decades), thus forming a generic property of memory.

### 659 **Limitations and future perspectives**

660         Our study presents limitations that should be addressed by future research. Due to the  
661 nature of the current investigation (a working memory task with long trials and subjectively  
662 determined variables), a relatively small number of unseen trials was acquired, thus making it

663 difficult to detect subtle effects. While our conclusions are supported by Bayes' Factor  
664 analyses, converging evidence from univariate and multivariate techniques, and similar results  
665 obtained with larger samples in the domain of activity-silent conscious working memory (e.g.,  
666 Rose et al., 2016; Wolff et al., 2017), a number of our observations are based on null effects,  
667 and it remains a possibility that we missed some target- and/or response-related activity on the  
668 unseen trials. Future research should thus aim at replicating the present findings with larger  
669 datasets or with more sensitive techniques, such as intra-cranial recordings. In particular, it  
670 might be interesting to further probe the relationship between seen, unseen correct, and  
671 unseen incorrect targets: A specific prediction of the proposed model is that unseen correct  
672 trials should possess enough activity to modify synaptic weights in high-level cortical circuits,  
673 yet without crossing the threshold for sustained activity and consciousness ("failed ignition").  
674 Unseen correct trials should thus share some of the processes that are found on seen trials and  
675 future research is necessary to directly test this hypothesis.

## 676 **Conclusion**

677 In contrast to a widely-held belief, our findings support the existence of genuine  
678 working memory in the absence of either conscious perception or sustained activity. Our  
679 proposal is that, following a transient encoding phase via active firing, non-conscious stimuli  
680 may be maintained by 'activity-silent' short-term changes in synaptic weights without any  
681 detectable neural activity, allowing above-chance retrieval for several seconds. Similar  
682 activity-silent codes also subserve conscious maintenance, though in this case periodic  
683 refreshing appears to stabilize the stored representations throughout the delay. Our findings  
684 thus highlight the need to refine our understanding of working memory, and to continuously  
685 challenge the limits of non-conscious processing.

## 686 **Methods**

### 687 **Subjects**

688 38 healthy volunteers participated in the present study (experiment 1:  $N = 17$ ,  $M_{\text{age}} =$   
689 23.3 years,  $SD_{\text{age}} = 2.8$  years, 10 men; experiment 2:  $N = 21$ ,  $M_{\text{age}} = 24.3$  years,  $SD_{\text{age}} = 3.8$   
690 years, 9 men). They gave written informed consent and received 80 or 15€ as compensation  
691 for the imaging and behavioral paradigms. Due to noisy recordings, only 13 of the 17 subjects  
692 in experiment 1 were retained for the MEG analyses. Although sample size had not  
693 specifically been estimated for our study, it thus was reasonable given typical experiments in  
694 the field.

### 695 **Experimental protocol**

696 Participants performed variations of a spatial delayed-response task, designed to assess  
697 retention of a target location under varying levels of subjective visibility (Figure 1A). Each  
698 trial began with the presentation of a central fixation cross (500ms), displayed in white ink on  
699 an otherwise black screen. In experiment 1, a faint gray target square (RGB: 89.25 89.25  
700 89.25) was flashed for 17ms in 1 out of 20 equally spaced, invisible positions along a circle  
701 centered on fixation (radius = 200 pixels; 8 repetitions/location). Another fixation cross  
702 (17ms) preceded the display of the mask (233ms). Mask elements were composed of four  
703 individual squares (two right above and below, and two to the left and right of the target  
704 stimulus), arranged to tightly surround the target square without overlapping it. They  
705 appeared simultaneously at all possible target locations. Mask contrast was adjusted on an  
706 individual basis in a separate calibration procedure (see below). A variable delay period with  
707 constant fixation followed the mask (experiment 1: 2.5, 3.0, 3.5, or 4.0s). On 50% of the trials  
708 in experiment 1, an unmasked distractor square, randomly placed and with the same duration  
709 as the target, was presented 1.5s into the delay period.

710 After the delay, 20 letters – drawn from a subset of lower-case letters of the alphabet  
711 (excluded: *e, j, n, p, t, v*) – were randomly presented in the 20 positions (2.5s). Participants  
712 were asked to identify the target location by speaking the name of the letter presented at the

713 location. They were instructed to always provide a response, guessing if necessary. A trial  
714 ended with the presentation of the word *Vu?* (French for *seen*) in the center of the screen  
715 (2.5s), cueing participants to rate the visibility of the target on the 4-point Perceptual  
716 Awareness Scale (PAS; 1: no experience of the target, 2: brief glimpse, 3: almost clear  
717 experience, 4: clear experience; Ramsøy and Overgaard, 2004) using the index, middle, ring,  
718 or little finger of their right hand (five-button non-magnetic response box, Cambridge  
719 Research Systems Ltd., Fiber Optic Response Pad). We instructed subjects to reserve a  
720 visibility rating of 1 for those trials, for which they had absolutely no perception of the target.  
721 The target square was also replaced by a blank screen on 20% of the trials, in order to obtain  
722 an objective measure of participants' sensitivity to the presence of the target. The inter-trial  
723 interval (ITI) lasted 1s. Subjects completed a total of 200 trials of this working memory task,  
724 divided into four separate experimental blocks. They also undertook two blocks of 100 trials  
725 each of a perception-only control paradigm, identical to the working memory task in all  
726 respects except that the delay period and target localization screen were omitted, such that the  
727 presentation of the mask immediately preceded subjects' visibility ratings. Task order  
728 (perception vs. working memory) was counterbalanced across participants.

729 Experiment 2 was designed to investigate the impact of a conscious working memory  
730 load on non-conscious working memory. Apart from the following exceptions, it was  
731 identical to experiment 1: A screen with either 1 (low load) or 5 (high load) centrally  
732 presented digits (1.5s) – randomly drawn (without replacement) from the numbers 1 through 9  
733 – as well as a 1s-fixation period were shown prior to the presentation of the target square.  
734 Following either a 0s- or a 4s-delay period, subjects first identified the target location by  
735 typing their responses on a standard AZERTY keyboard (4s). The French word for *numbers*  
736 (*Numéros?*) then probed participants to recall the sequence of digits in the correct order.  
737 Responses were again logged on the keyboard during a period of 4.5s. Subjects last rated

738 target visibility as in experiment 1 (3s). The ITI varied between 1 and 2s. Participants  
739 completed two experimental blocks of 100 trials each.

#### 740 **Calibration task**

741 Prior to the experimental tasks, each participant's perceptual threshold was estimated  
742 in order to ensure roughly equal proportions of seen and unseen trials. Subjects completed 150  
743 (experiment 1: 3 blocks) or 125 (experiment 2: 5 blocks) trials of a modified version of the  
744 working memory task (no distractor, delay duration: 2s in experiment 1 and 0s in experiment  
745 2), during which mask contrast was either increased (following a visibility rating of 2, 3, or 4)  
746 or decreased (following a visibility rating of 1) on each target-present trial according to a  
747 double-staircase procedure. Individual perceptual thresholds to be used in the main tasks were  
748 derived by averaging the mask contrasts from the last four switches from seen to unseen (or  
749 vice versa) in each staircase.

#### 750 **Behavioral analyses**

751 We analyzed our behavioral data in Matlab R2014a (MathWorks Inc., Natick, MA;  
752 code available upon request) and SPSS Statistics Version 20.0 (IBM, Armonk, NY), using  
753 repeated-measures analyses of variance (ANOVAs). Only meaningful trials without missing  
754 responses were included in any analysis. Distributions of localization responses were  
755 computed for visibility categories with at least five trials per subject. Objective working  
756 memory performance was quantified via two complementary measures. The *rate of correct*  
757 *responding* was defined as the proportion of trials within two positions (i.e., +/- 36°) of the  
758 actual target location and served as an index of the amount of information that could be  
759 retained. Because 5 out of 20 locations were counted as correct, chance on this measure was  
760 25%. The *precision* of working memory was estimated as the dispersion (standard deviation)  
761 of spatial responses. In particular, we modeled the observed distribution of responses  $D(n)$  as

762 a mixture of a uniform distribution (random guessing) and an unknown probability  
763 distribution  $d$  (“true working memory”):

764

$$765 \quad (1) D(n) = \frac{p}{N} + (1-p)d(n)$$

766

767 where  $p$  refers to the probability that a given trial is responded to using random guessing;  $N$   
768 to the number of target locations ( $N = 20$ ); and  $n$  is the deviation from the true target location.  
769 We assumed that  $d(n) = 0$  for deviations beyond a fixed limit  $a$  (with  $a = 2$ ). This hypothesis  
770 allowed us to estimate  $p$  from the mean of that part of the distribution  $D$  for which one may  
771 safely assume no contribution of working memory:

772

$$773 \quad (2) \hat{p} = \frac{\sum D(n) | n \text{ outside } [-a, a]}{(N-2a-1)} * N$$

774

775 where the model is designed in such a way as to ensure that  $\hat{p} = 1$  if  $D$  is a uniform  
776 distribution (i.e., 100% of random guessing) and  $\hat{p} = 0$  if  $D$  vanishes outside the region of  
777 correct responding (i.e., 0% of random guessing). There needs to be at least chance  
778 performance inside the region of correct responding, so

779

$$780 \quad (3) \sum D(n) | n \in [-a, a] \geq \frac{2a-1}{N}$$

781

782 which ensures  $0 \leq \hat{p} \leq 1$ . This is the reason why, when computing precision, we included  
783 only subjects whose rate of correct responding for unseen trials, collapsed across all  
784 experimental conditions, significantly exceeded chance performance (i.e., 25%) in a  $\chi^2$ -test ( $p$   
785  $< .05$ ). An estimate of  $d$ ,  $\hat{d}$ , can then be derived in two steps from Equation 1 as



786

787 
$$(4) \delta(n) = \frac{D(n) - \frac{\hat{p}}{N}}{1 - \hat{p}}$$

788 
$$(5) \hat{d}(n) = \frac{\delta(n)|_{n \in [-a, a]}}{\sum \delta(n)|_{n \in [-a, a]}}$$

789

790 We note that the distribution  $\delta$  has residual, yet negligible, positive and negative mass (due to  
791 noise) outside the region of correct responding. In order to obtain  $\hat{d}$ , we therefore restricted  
792 the distribution  $\delta$  to  $[-a, a]$ , set all negative values to 0, and renormalized its mass to 1. The  
793 precision of the representation of the target location in working memory was then defined as  
794 the standard deviation of that distribution.

### 795 **MEG recordings and preprocessing**

796 In experiment 1, we recorded MEG with a 306-channel (102 sensor triplets: 1  
797 magnetometer and 2 orthogonal planar gradiometers), whole-head setup by  
798 ElektaNeuromag® (Helsinki, Finland) at 1000Hz with a hardware bandpass filter between 0.1  
799 and 330Hz. Eye movements as well as heart rate were monitored with vertical and horizontal  
800 EOG and ECG channels. Prior to installation of the subject in the MEG chamber, we digitized  
801 three head landmarks (nasion and pre-auricular points), four head position indicator (HPI)  
802 coils placed over frontal and mastoïdian skull areas, and 60 additional locations outlining the  
803 participant's head with a 3-dimensional Fastrak system (Polhemus, USA). Head position was  
804 measured at the beginning of each run.

805 Our preprocessing pipeline followed Marti et al. (2015). Using MaxFilter Software  
806 (ElektaNeuromag®, Helsinki, Finland), raw MEG signals were first cleaned of head  
807 movements, bad channels, and magnetic interference originating from outside the MEG  
808 helmet (Taulu et al., 2004), and then downsampled to 250Hz. We conducted all further  
809 preprocessing steps with the Fieldtrip toolbox (<http://www.fieldtriptoolbox.org/>; Oostenveld

810 et al., 2011) run in a Matlab R2014a environment. Initially, MEG data were epoched between  
811 -0.5 and +2.5s with respect to target onset for all stimulus-locked, and between -0.5 and +0.8s  
812 with respect to the onset of the response screen for all response-locked analyses. Trials  
813 contaminated by muscle or other movement artifacts were then identified and rejected in a  
814 semi-automated procedure, for which the variance of the MEG signals across sensors served  
815 as an index of contamination. To remove any residual eye-movement and cardiac artifacts, we  
816 performed independent component analysis separately for each channel type, visually  
817 inspected the topographies and time courses of the first 30 components, and subtracted any  
818 contaminated component from the MEG data. Except for analyses requiring higher spatial  
819 precision (i.e., circular-linear correlations and decoding), results are presented for  
820 magnetometers only.

821 Further preprocessing steps depended on the nature of the subsequent analysis: Epochs  
822 retained for investigations based on evoked responses (i.e., ERFs, decoding, circular-linear  
823 correlations) were low-pass filtered at 30Hz, while time-frequency decompositions relied on  
824 entirely unfiltered data. In the latter case, a sliding, frequency-independent Hann taper  
825 (window size: 500ms, step size: 20ms) was convolved with the unfiltered epochs in order to  
826 extract an estimate of power between 1 and 99Hz (in 2Hz steps) to identify the neural  
827 correlates of conscious and non-conscious perception and working memory in the frequency  
828 domain. Prior to univariate or multivariate statistical analysis, data (ERFs, time-frequency  
829 power estimates) were baseline corrected using a period between -200 and -50ms.

### 830 **Circular-linear correlations**

831 To localize and track the neural representations of target, response, and distractor  
832 location, filtered epochs were transformed into circular-linear correlation coefficients.  
833 Following King et al. (2016), we combined the two linear correlation coefficients between the  
834 MEG signal and the sine and cosine of the angle defining the location in question (i.e., target,

835 distractor, or response). An empirical null distribution was generated for each condition  
836 separately by shuffling the labels (i.e., target, distractor, or response location) at the  
837 corresponding time points and averaging the resulting distribution from 1000 such  
838 permutations.

839         Due to the spatial nature of our task, there is a possibility that subjects could have  
840 systematically moved their eyes after the presentation of the target, thus contaminating the  
841 correlation analyses. However, several lines of evidence suggest that this was not the case:  
842 First, participants were carefully instructed not to move their eyes. A close inspection of the  
843 EOG traces confirmed that subjects successfully implemented this request and did not display  
844 any strategic eye movements. Second, we carefully removed any trials contaminated by such  
845 movements as part of our preprocessing procedure. Third, the topographical patterns of the  
846 correlations show that the signal primarily originated in occipital and parietal channels. Eye  
847 movements therefore unlikely have driven the circular-linear correlations.

#### 848 **Sources**

849         Individual anatomical magnetic resonance images (MRI), obtained with a 3D T1-  
850 weighted spoiled gradient recalled pulse sequence (voxel size: 1 \* 1 \* 1.1mm; repetition time  
851 [TR]: 2,300ms; echo time [TE]: 2.98ms; field of view [FOV]: 256 \* 240 \* 176mm; 160  
852 slices) in a 3T Tim Trio Siemens scanner, were first segmented into gray/white matter as well  
853 as subcortical structures with FreeSurfer (<https://surfer.nmr.mgh.harvard.edu/>). We then  
854 reconstructed the cortical, scalp, and head surfaces in Brainstorm  
855 (<http://neuroimage.usc.edu/brainstorm>; Tadel et al., 2011) and co-registered these anatomical  
856 images with the MEG signals, using the HPI coils and the digitized head shape as a reference.  
857 Current density distributions on the cortical surface were subsequently estimated separately  
858 for each condition and subject. Specifically, we employed an analytical model with  
859 overlapping spheres to compute the leadfield matrix and modeled neuronal current sources

860 with an unconstrained (dipole orientation loosening factor: 0.2) weighted minimum-norm  
861 current estimate (wMNE; depth-weighting factor: 0.5) and a noise covariance obtained from  
862 the baseline period of all trials. Average time-frequency power in the alpha (8 – 12Hz) and  
863 beta (13 – 30 Hz) band was then estimated with complex Morlet wavelets using the  
864 Brainstorm default parameters, the resulting transformations projected onto the ICBM 152  
865 anatomical template (Fonov et al., 2011, 2009), and the contrasts between the conditions of  
866 interest computed. Group averages for spatial clusters of at least 150 vertices are shown in dB  
867 relative to baseline and were thresholded at 60% of the maximum amplitude (cortex smoothed  
868 at 60%).

### 869 **Multivariate pattern analyses**

870 We employed the Scikit-Learn package (Pedregosa et al., 2011) as implemented in  
871 MNE 0.13 (Gramfort, 2013; Gramfort et al., 2014) in order to conduct our multivariate  
872 pattern analyses (MVPA). Following Marti et al. (2015) and King et al. (2016), we fit linear  
873 estimators at each time sample within each participant to isolate the topographical patterns  
874 best differentiating our experimental conditions. Support vector machines (Chang and Lin,  
875 2011) were trained in the case of categorical data (i.e., visibility/accuracy) and a combination  
876 of two linear support vector regressions was used for circular data (i.e., target/response  
877 location) to estimate an angle from the arctangent of the separately predicted sine and cosine  
878 of the labels of interest.

879 A 5- (for categorical variables) or, due to the much larger number of labels, 2-fold (for  
880 circular variables), stratified cross-validation procedure was used in order to avoid overfitting:  
881 MEG data were first split into five (two) sets of trials with the same proportion of samples for  
882 each class. Within each fold, four (one) of these sets served as the training data and the  
883 remainder as the testing data. Model fitting, including all preprocessing steps, was exclusively  
884 performed on the training set. 50% of the most informative features (i.e., channels) were

885 selected by means of a simple, univariate analysis of variance to reduce the dimensionality of  
886 the data (Charles et al., 2014; Haynes and Rees, 2006), the remaining channel-time features  $z$ -  
887 score normalized, and a weighting procedure applied in order to counteract the effects of any  
888 class imbalances. The classifier was then trained on the resulting data and applied to the left-  
889 out trials in order to identify the hyperplane (i.e., topography) best suited to separate the  
890 classes. This sequence of events (univariate feature selection, normalization, training and  
891 testing) was repeated five (two) times, ensuring that each trial would be included in the test  
892 set once.

893         Within the same cross-validation loop, we also evaluated the ability of each classifier  
894 to discriminate the experimental conditions of interest at all other time samples (i.e.,  
895 generalization across time). This kind of MVPA results in a temporal generalization matrix, in  
896 which each entry represents the decoding performance of each classifier trained at time point  $t$   
897 and tested at time point  $t'$ , and in which the diagonal corresponds to classifiers trained and  
898 tested on the same time points (King and Dehaene, 2014). Importantly, when interrogating the  
899 capacity of our classifiers to generalize across tasks or labels (e.g., from the perception to the  
900 working memory task, or from seen to unseen correct target locations), we modified the  
901 aforementioned cross-validation procedure to capitalize on the independence of our training  
902 and testing data (see  
903 [http://martinos.org/mne/dev/auto\\_examples/decoding/plot\\_decoding\\_time\\_generalization\\_con](http://martinos.org/mne/dev/auto_examples/decoding/plot_decoding_time_generalization_conditions.html#example-decoding-plot-decoding-time-generalization-conditions-py)  
904 [ditions.html#example-decoding-plot-decoding-time-generalization-conditions-py](http://martinos.org/mne/dev/auto_examples/decoding/plot_decoding_time_generalization_conditions.html#example-decoding-plot-decoding-time-generalization-conditions-py)). As such,  
905 classifiers from each training set were directly applied to the entire testing set and the  
906 respective predictions averaged.

907         Classifiers for categorical data generated a continuous output in the form of the  
908 distance between the respective sample and the separating hyperplane for each test trial. In  
909 order to be able to compare classification performance across subjects, we then applied a

910 receiver operating characteristic analysis across trials within each participant and summarized  
911 overall effect sizes with the area under the curve (AUC). Unlike average decoding accuracy,  
912 the AUC serves as an unbiased measure of decoding performance as it represents the true-  
913 positive rate (e.g., a trial was correctly categorized as seen) as a function of the false-positive  
914 rate (e.g., a trial was incorrectly categorized as seen). Chance performance, corresponding to  
915 equal proportions of true and false positives, therefore leads to an AUC of 0.5. Any value  
916 greater than this critical level implies better-than-chance performance, with an AUC of 1  
917 indicating a perfect prediction for any given class. In contrast, classifiers for circular data  
918 were first summarized by computing the mean absolute difference between the predicted and  
919 the actual angle (range: 0 to  $\pi$ ; chance:  $\pi/2$ ) and then transformed into an “accuracy” score  
920 (range:  $-\pi/2$  to  $\pi/2$ ; chance: 0). To facilitate comparability between different conditions, an  
921 additional baseline-correction was then performed.

## 922 **Statistical analyses**

923 We performed statistical analyses across subjects. For the ERF and time-frequency  
924 data, cluster-based, non-parametric *t*-tests with Monte Carlo permutations were used to  
925 identify significant differences between experimental conditions (Maris and Oostenveld,  
926 2007). Further planned comparisons of ERF time courses (seen vs. unseen) in a-priori defined  
927 spatio-temporal regions of interest (i.e., P3b time window: 300 – 600ms) were conducted with  
928 non-parametric signed-rank tests ( $p_{\text{uncorrected}} < .05$ ). A correction for multiple comparisons was  
929 then applied with a false discovery rate ( $p_{\text{FDR}} < .05$ ).

930 Non-parametric signed-rank tests ( $p_{\text{uncorrected}} < .05$ ) were also employed to evaluate  
931 decoding performance and the strength of circular-linear correlations. Specifically, we  
932 assessed whether classifiers could predict the trials’ classes better than chance (categorical  
933 data:  $\text{AUC} > 0.5$ ; circular data:  $\text{rad} > 0$ ) and whether circular-linear correlation coefficients  
934 deviated from an empirical baseline ( $\Delta \rho > 0$ ). We report temporal averages over four a-

935 priori time bins, corresponding to an early perceptual period (0.1 – 0.3s), the P3b time  
 936 window (0.3 – 0.6s), and the first (0.6 – 1.55s) and second (1.55 – 2.53s) part of the delay  
 937 period. To capitalize on the increased spatial selectivity of gradiometers, averaged time  
 938 courses of these two channels are shown for circular-linear correlations.

939 Bayesian statistics, based on either two- (time-frequency analyses) or one-sided  
 940 (circular-linear correlations)  $t$ -tests, were also computed when appropriate with a scale factor  
 941 of  $r = .707$  (Rouder et al., 2009).

## 942 **Simulations**

943 A one-dimensional, recurrent continuous attractor neural network (CANN) model  
 944 (Mongillo et al., 2008) was adapted in order to simulate the experimental findings (Figure  
 945 7A). Individual neurons were aligned according to their preferred stimulus value, enabling the  
 946 network to encode angular position of a target stimulus (range:  $-\pi$  to  $\pi$ ; periodic boundary  
 947 condition). The dynamics of this system were determined by the synaptic currents of each  
 948 neuron given by

949

$$950 \quad (6) \quad \tau \frac{\partial h_E(\theta, t)}{\partial t} = -h_0 + \rho \int_{-\pi}^{\pi} J(\theta, \theta') u(\theta', t) x(\theta', t) R_E(\theta', t) d\theta' - J_{EI} R_I + I_b + \delta_1 \xi_1(\theta, t) + I_e + \delta_2 \xi_2(\theta, t),$$

951

$$(7) \quad \frac{\partial u(\theta, t)}{\partial t} = \frac{U - u(\theta, t)}{\tau_f} + U [1 - u(\theta, t)] R_E(\theta, t),$$

952

$$(8) \quad \frac{\partial x(\theta, t)}{\partial t} = \frac{1 - x(\theta, t)}{\tau_d} - u(\theta, t) x(\theta, t) R_E(\theta, t), \text{ and}$$

953

$$(9) \quad \tau \frac{\partial h_I}{\partial t} = -h_I + J_{IE} \int_{-\pi}^{\pi} R_E(\theta, t),$$

954

955 where  $\tau$  describes the time constant of firing rate dynamics (in the order of milliseconds);  $\rho$   
 956 refers to neuronal density;  $h_E(\theta, t)$  and  $R_E(\theta, t)$  capture the synaptic current to and firing rate  
 957 of neurons with preference  $\theta$  at time  $t$  respectively; and  $R(h) = \alpha \ln(1 + \exp(h/\alpha))$  is the neural  
 958 gain chosen in the form of a smoothed threshold-linear function.  $J_{IE}$  and  $J_{EI}$  represent the

959 connection strength between excitatory and inhibitory neurons. All excitatory neurons  
 960 received a constant background input,  $I_e$ , reflecting the arousal signal when the neural system  
 961 was engaged in a working memory task.  $\delta_I \zeta_I$  is background noise;  $I_e$ , any external stimulus  
 962 (e.g., target, mask, and recall signal); and  $\delta_I \zeta_I(t)$  the noise related to those external stimuli.  $u$   
 963 ( $\theta, t$ ) and  $x(\theta, t)$  denote the short-term synaptic facilitation (STF) and depression (STD)  
 964 effects at time  $t$  of neurons with preference  $\theta$ , respectively. The short-term plasticity  
 965 dynamics are characterized by the following parameters:  $J_I$  (absolute efficacy),  $U$  (increment  
 966 of the release probability when a spike arrives),  $\tau_f$  and  $\tau_d$  (facilitation and depression time  
 967 constants). The STF value  $u(\theta, t)$  is facilitated whenever a spike arrives, and decays to the  
 968 baseline  $U$  within the time  $\tau_f$ . The neurotransmitter value  $x(\theta, t)$  is utilized by each spike in  
 969 proportion to  $u(\theta, t)$  and then recovers to its baseline, 1, within the time  $\tau_d$ .

970  $J(\theta, \theta')$  is the interaction strength from neurons at  $\theta$  to neurons at  $\theta'$  and is chosen to  
 971 be

972

$$(10) J(\theta, \theta') = \begin{cases} J_1 \cos[B * (\theta - \theta')] - J_0, & \text{if } B * (\theta - \theta') \in [-\arccos(-J_0/J_1), \arccos(-J_0/J_1)], \\ -J_0, & \text{else} \end{cases}$$

973 where  $J_0$ ,  $J_1$ , and  $B$  are constants which determine the connection strength between the  
 974 neurons. Note that  $J(\theta, \theta')$  is a function of  $\theta - \theta'$ , i. e., the neuronal interactions are  
 975 translation-invariant in the space of neural preferred stimuli. The other parameters of the  
 976 system were as follows:  $\tau = 0.008\text{s}$ ,  $\tau_f = 4\text{s}$ ,  $\tau_d = 0.3\text{s}$ ,  $J_I = 12$ ,  $J_0 = 1$ ,  $J_{EI} = 1.9$ ,  $J_{IE} = 1.8$ ,  $I_b = -$   
 977  $0.1\text{Hz}$ ,  $\delta_I = 0.3$ ,  $\delta_2 = 9$ ,  $N = 100$ ,  $\alpha = 1.5$ ,  $B = 2.2$ .

978 During our simulations, we first presented a target signal with an amplitude of  $A_{\text{target}} =$   
 979  $390\text{Hz}$  at a random location (50ms), waited for 17ms, and then applied a mask signal to all the  
 980 neurons in the system (200ms). The amplitude of the mask signal was initially varied in order



981 to determine a critical value which would produce two distinct maintenance patterns, but was  
982 then fixed at a threshold of  $A_{\text{mask}} = 62\text{Hz}$ . At the end of a 3s-delay period, a non-specific recall  
983 signal was given for 50ms with  $A_{\text{recall}} = 10\text{Hz}$ . Remembered target position was calculated as  
984 the population vector angle during this time period.

### 985 **Figure Supplements**

986 Figure 2 – Supplement 1. Perceptual sensitivity does not correlate with working memory  
987 performance on unseen trials.

988 Figure 4 – Supplement 1. Alpha- and beta-band desynchronizations serve as a general  
989 signature of conscious processing and conscious working memory.

990 Figure 4 – Supplement 2. Seen and unseen correct trials do not share the same discriminative  
991 decoding axis.

992 Figure 4 – Supplement 3. Bayesian statistics for the time-frequency analyses.

993 Figure 5 – Supplement 1. Representation of seen target locations during conscious perception  
994 and working memory.

995 Figure 5 – Supplement 2. Circular-linear correlations and multivariate decoding reveal similar  
996 time courses for target location.

997 Figure 5 – Supplement 3. Tracking target/response location on unseen correct and incorrect  
998 trials with multivariate decoding.

999 Figure 6 – Supplement 1. Topographies for circular-linear correlations with response location  
1000 as a function of visibility.

1001 Figure 6 – Supplement 2. Circular-linear correlations and multivariate decoding reveal similar  
1002 time courses for response location.

### 1003 **Acknowledgements**

1004 We gratefully acknowledge Henrik Ueberschär, Leila Azizi, and Virginie Van  
1005 Wassenhove for their invaluable daily support and stimulating discussion.

1006

## Competing Interests

1007 The authors declare no competing interests.

1008

## References

- 1009 Baars, B.J., Franklin, S., 2003. How conscious experience and working memory interact.  
1010 Trends Cogn. Sci. 7, 166–172. doi:10.1016/S1364-6613(03)00056-1
- 1011 Baddeley, A., 2003. Working memory: looking back and looking forward. Nat. Rev.  
1012 Neurosci. 4, 829–839. doi:10.1038/nrn1201
- 1013 Bergström, F., Eriksson, J., 2014. Maintenance of non-consciously presented information  
1014 engages the prefrontal cortex. Front. Hum. Neurosci. 8.  
1015 doi:10.3389/fnhum.2014.00938
- 1016 Bergström, F., Eriksson, J., 2015. The conjunction of non-consciously perceived object  
1017 identity and spatial position can be retained during a visual short-term memory task.  
1018 Front. Psychol. 6. doi:10.3389/fpsyg.2015.01470
- 1019 Chang, C.-C., Lin, C.-J., 2011. LIBSVM: A library for support vector machines. ACM Trans.  
1020 Intell. Syst. Technol. 2, 1–27. doi:10.1145/1961189.1961199
- 1021 Charles, L., King, J.-R., Dehaene, S., 2014. Decoding the Dynamics of Action, Intention, and  
1022 Error Detection for Conscious and Subliminal Stimuli. J. Neurosci. 34, 1158–1170.  
1023 doi:10.1523/JNEUROSCI.2465-13.2014
- 1024 Chaudhuri, R., Bernacchia, A., Wang, X.-J., 2014. A diversity of localized timescales in  
1025 network activity. eLife 3. doi:10.7554/eLife.01239
- 1026 Chong, T.T.-J., Husain, M., Rosenthal, C.R., 2014. Recognizing the unconscious. Curr. Biol.  
1027 24, R1033–R1035. doi:10.1016/j.cub.2014.09.035
- 1028 Cooper, N.R., Croft, R.J., Dominey, S.J., Burgess, A.P., Gruzelier, J.H., 2003. Paradox lost?  
1029 Exploring the role of alpha oscillations during externally vs. internally directed  
1030 attention and the implications for idling and inhibition hypotheses. Int. J.  
1031 Psychophysiol. 47, 65–74. doi:10.1016/S0167-8760(02)00107-1
- 1032 Dehaene, S., Changeux, J.-P., 2011. Experimental and Theoretical Approaches to Conscious  
1033 Processing. Neuron 70, 200–227. doi:10.1016/j.neuron.2011.03.018
- 1034 Dehaene, S., Charles, L., King, J.-R., Marti, S., 2014. Toward a computational theory of  
1035 conscious processing. Curr. Opin. Neurobiol. 25, 76–84.  
1036 doi:10.1016/j.conb.2013.12.005
- 1037 Dehaene, S., Naccache, L., 2001. Towards a cognitive neuroscience of consciousness: basic  
1038 evidence and a workspace framework. Cognition 79, 1–37.
- 1039 Del Cul, A., Baillet, S., Dehaene, S., 2007. Brain Dynamics Underlying the Nonlinear  
1040 Threshold for Access to Consciousness. PLoS Biol. 5, e260.  
1041 doi:10.1371/journal.pbio.0050260
- 1042 Dupoux, E., Gardelle, V. de, Kouider, S., 2008. Subliminal speech perception and auditory  
1043 streaming. Cognition 109, 267–273. doi:10.1016/j.cognition.2008.06.012
- 1044 Dutta, A., Shah, K., Silvanto, J., Soto, D., 2014. Neural basis of non-conscious visual working  
1045 memory. NeuroImage 91, 336–343. doi:10.1016/j.neuroimage.2014.01.016
- 1046 Eriksson, J., Vogel, E.K., Lansner, A., Bergström, F., Nyberg, L., 2015. Neurocognitive  
1047 Architecture of Working Memory. Neuron 88, 33–46.  
1048 doi:10.1016/j.neuron.2015.09.020
- 1049 Fonov, V., Evans, A.C., Botteron, K., Almlí, C.R., McKinstry, R.C., Collins, D.L., 2011.  
1050 Unbiased average age-appropriate atlases for pediatric studies. NeuroImage 54, 313–  
1051 327. doi:10.1016/j.neuroimage.2010.07.033

1052 Fonov, V., Evans, A., McKinstry, R., Almlí, C., Collins, D., 2009. Unbiased nonlinear  
1053 average age-appropriate brain templates from birth to adulthood. *NeuroImage* 47,  
1054 S102. doi:10.1016/S1053-8119(09)70884-5

1055 Fuentemilla, L., Penny, W.D., Cashdollar, N., Bunzeck, N., Düzel, E., 2010. Theta-Coupled  
1056 Periodic Replay in Working Memory. *Curr. Biol.* 20, 606–612.  
1057 doi:10.1016/j.cub.2010.01.057

1058 Fujisawa, S., Amarasingham, A., Harrison, M.T., Buzsáki, G., 2008. Behavior-dependent  
1059 short-term assembly dynamics in the medial prefrontal cortex. *Nat. Neurosci.* 11, 823–  
1060 833. doi:10.1038/nn.2134

1061 Gaillard, R., Dehaene, S., Adam, C., Clémenceau, S., Hasboun, D., Baulac, M., Cohen, L.,  
1062 Naccache, L., 2009. Converging Intracranial Markers of Conscious Access. *PLoS*  
1063 *Biol.* 7, e1000061. doi:10.1371/journal.pbio.1000061

1064 Gramfort, A., 2013. MEG and EEG data analysis with MNE-Python. *Front. Neurosci.* 7.  
1065 doi:10.3389/fnins.2013.00267

1066 Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C.,  
1067 Parkkonen, L., Hämäläinen, M.S., 2014. MNE software for processing MEG and EEG  
1068 data. *NeuroImage* 86, 446–460. doi:10.1016/j.neuroimage.2013.10.027

1069 Greenwald, A.G., Draine, S.C., Abrams, R.L., 1996. Three cognitive markers of unconscious  
1070 semantic activation. *Science* 273, 1699–1702.

1071 Gross, J., Schnitzler, A., Timmermann, L., Ploner, M., 2007. Gamma Oscillations in Human  
1072 Primary Somatosensory Cortex Reflect Pain Perception. *PLoS Biol.* 5, e133.  
1073 doi:10.1371/journal.pbio.0050133

1074 Harrison, S.A., Tong, F., 2009. Decoding reveals the contents of visual working memory in  
1075 early visual areas. *Nature* 458, 632–635. doi:10.1038/nature07832

1076 Haynes, J.-D., Rees, G., 2006. Decoding mental states from brain activity in humans. *Nat.*  
1077 *Rev. Neurosci.* 7, 523–534. doi:10.1038/nrn1931

1078 Jensen, O., Mazaheri, A., 2010. Shaping Functional Architecture by Oscillatory Alpha  
1079 Activity: Gating by Inhibition. *Front. Hum. Neurosci.* 4.  
1080 doi:10.3389/fnhum.2010.00186

1081 King, J.-R., Pescetelli, N., Dehaene, S., 2016. Brain mechanisms underlying the brief  
1082 maintenance of seen and unseen sensory information. *Neuron.* 92, 1122–1134. doi:  
1083 10.1016/j.neuron.2016.10.051

1084 King, J.-R., Dehaene, S., 2014. Characterizing the dynamics of mental representations: the  
1085 temporal generalization method. *Trends Cogn. Sci.* 18, 203–210.  
1086 doi:10.1016/j.tics.2014.01.002

1087 King, J.-R., Gramfort, A., Schurger, A., Naccache, L., Dehaene, S., 2014. Two Distinct  
1088 Dynamic Modes Subtend the Detection of Unexpected Sounds. *PLoS ONE* 9, e85791.  
1089 doi:10.1371/journal.pone.0085791

1090 Lamme, V.A.F., Roelfsema, P.R., 2000. The distinct modes of vision offered by feedforward  
1091 and recurrent processing. *Trends Neurosci.* 23, 571–579. doi:10.1016/S0166-  
1092 2236(00)01657-X

1093 Lundqvist, M., Herman, P., Lansner, A., 2011. Theta and gamma power increases and  
1094 alpha/beta power decreases with memory load in an attractor network model. *J. Cogn.*  
1095 *Neurosci.* 23, 3008–3020. doi:10.1162/jocn\_a\_00029

1096 Lundqvist, M., Rose, J., Herman, P., Brincat, S.L., Buschman, T.J., Miller, E.K., 2016.  
1097 Gamma and Beta Bursts Underlie Working Memory. *Neuron* 90, 152–164.  
1098 doi:10.1016/j.neuron.2016.02.028

1099 Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. *J.*  
1100 *Neurosci. Methods* 164, 177–190. doi:10.1016/j.jneumeth.2007.03.024

- 1101 Marti, S., King, J.-R., Dehaene, S., 2015. Time-Resolved Decoding of Two Processing  
1102 Chains during Dual-Task Interference. *Neuron* 88, 1297–1307.  
1103 doi:10.1016/j.neuron.2015.10.040
- 1104 Merikle, P., 2001. Perception without awareness: perspectives from cognitive psychology.  
1105 *Cognition* 79, 115–134. doi:10.1016/S0010-0277(00)00126-8
- 1106 Mongillo, G., Barak, O., Tsodyks, M., 2008. Synaptic Theory of Working Memory. *Science*  
1107 319, 1543–1546. doi:10.1126/science.1150769
- 1108 Naghavi, H.R., Nyberg, L., 2005. Common fronto-parietal activity in attention, memory, and  
1109 consciousness: Shared demands on integration? *Conscious. Cogn.* 14, 390–425.  
1110 doi:10.1016/j.concog.2004.10.003
- 1111 Oberauer, K., 2002. Access to information in working memory: Exploring the focus of  
1112 attention. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 411–421. doi:10.1037//0278-  
1113 7393.28.3.411
- 1114 Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.-M., 2011. FieldTrip: Open Source Software  
1115 for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data.  
1116 *Comput. Intell. Neurosci.* 2011, 1–9. doi:10.1155/2011/156869
- 1117 Pedregosa, Fabian, Varoquaux, Gael, Gramfort, Alexandre, Michel, Vincent, Thirion,  
1118 Bertrand, Grisel, Olivier, Blondel, Mathieu, Prettenhofer, Peter, Weiss, Ron, Dubourg,  
1119 Vincent, Vanderplas, Jake, Passos, Alexandre, Cournapeau, David, Brucher,  
1120 Matthieu, Perrot, Matthieu, Duchesnay, Edouard, 2011. Scikit-learn: Machine  
1121 Learning in Python. *J. Mach. Learn. Res.* 2825–2830.
- 1122 Ramsøy, T.Z., Overgaard, M., 2004. Introspection and subliminal perception. *Phenomenol.*  
1123 *Cogn. Sci.* 3, 1–23. doi:10.1023/B:PHEN.0000041900.30172.e8
- 1124 Roelfsema, P.R., 2015. The role of the different layers of primary visual cortex in working  
1125 memory. *J. Vis.* 15, 1406. doi:10.1167/15.12.1406
- 1126 Rose, N.S., LaRocque, J.J., Riggall, A.C., Gosseries, O., Starrett, M.J., Meyering, E.E.,  
1127 Postle, B.R., 2016. Reactivation of latent working memories with transcranial  
1128 magnetic stimulation. *Science* 354, 1136–1139. doi:10.1126/science.aah7011
- 1129 Rosenthal, C.R., Andrews, S.K., Antoniadis, C.A., Kennard, C., Soto, D., 2016. Learning and  
1130 Recognition of a Non-conscious Sequence of Events in Human Primary Visual  
1131 Cortex. *Curr. Biol.* 26, 834–841. doi:10.1016/j.cub.2016.01.040
- 1132 Rouder, J.N., Speckman, P.L., Sun, D., Morey, R.D., Iverson, G., 2009. Bayesian t tests for  
1133 accepting and rejecting the null hypothesis. *Psychon. Bull. Rev.* 16, 225–237.  
1134 doi:10.3758/PBR.16.2.225
- 1135 Salti, M., Monto, S., Charles, L., King, J.-R., Parkkonen, L., Dehaene, S., 2015. Distinct  
1136 cortical codes and temporal dynamics for conscious and unconscious percepts. *eLife* 4.  
1137 doi:10.7554/eLife.05652
- 1138 Schurger, A., Sarigiannidis, I., Naccache, L., Sitt, J.D., Dehaene, S., 2015. Cortical activity is  
1139 more stable when sensory stimuli are consciously perceived. *Proc. Natl. Acad. Sci.*  
1140 112, E2083–E2092. doi:10.1073/pnas.1418730112
- 1141 Sergent, C., Baillet, S., Dehaene, S., 2005. Timing of the brain events underlying access to  
1142 consciousness during the attentional blink. *Nat. Neurosci.* 8, 1391–1400.  
1143 doi:10.1038/nn1549
- 1144 Soto, D., Mäntylä, T., Silvanto, J., 2011. Working memory without consciousness. *Curr. Biol.*  
1145 21, R912–R913. doi:10.1016/j.cub.2011.09.049
- 1146 Soto, D., Silvanto, J., 2014. Reappraising the relationship between working memory and  
1147 conscious awareness. *Trends Cogn. Sci.* 18, 520–525. doi:10.1016/j.tics.2014.06.005
- 1148 Sprague, T.C., Ester, E.F., Serences, J.T., 2016. Restoring Latent Visual Working Memory  
1149 Representations in Human Cortex. *Neuron* 91, 694–707.  
1150 doi:10.1016/j.neuron.2016.07.006

- 1151 Stein, T., Kaiser, D., Hesselmann, G., 2016. Can working memory be non-conscious?  
1152 Neurosci. Conscious. 2016, niv011. doi:10.1093/nc/niv011
- 1153 Stokes, M.G., 2015. “Activity-silent” working memory in prefrontal cortex: a dynamic coding  
1154 framework. Trends Cogn. Sci. 19, 394–405. doi:10.1016/j.tics.2015.05.004
- 1155 Stokes, M.G., Wolff, M.J., Spaak, E., 2015. Decoding Rich Spatial Information with High  
1156 Temporal Resolution. Trends Cogn. Sci. 19, 636–638. doi:10.1016/j.tics.2015.08.016
- 1157 Tadel, F., Baillet, S., Mosher, J.C., Pantazis, D., Leahy, R.M., 2011. Brainstorm: A User-  
1158 Friendly Application for MEG/EEG Analysis. Comput. Intell. Neurosci. 2011, 1–13.  
1159 doi:10.1155/2011/879716
- 1160 Taulu, S., Kajola, M., Simola, J., 2004. Suppression of interference and artifacts by the Signal  
1161 Space Separation Method. Brain Topogr. 16, 269–275.
- 1162 Tononi, G., Koch, C., 2008. The Neural Correlates of Consciousness: An Update. Ann. N. Y.  
1163 Acad. Sci. 1124, 239–261. doi:10.1196/annals.1440.004
- 1164 van Gaal, S., Lamme, V.A.F., 2012. Unconscious High-Level Information Processing:  
1165 Implication for Neurobiological Theories of Consciousness. The Neuroscientist 18,  
1166 287–301. doi:10.1177/1073858411404079
- 1167 Wolff, M.J., Ding, J., Myers, N.E., Stokes, M.G., 2015. Revealing hidden states in visual  
1168 working memory using electroencephalography. Front. Syst. Neurosci. 9.  
1169 doi:10.3389/fnsys.2015.00123
- 1170 Wolff, M.J., Jochim, J., Akyürek, E.G., Stokes, M.G., 2017. Dynamic hidden states  
1171 underlying working-memory-guided behavior. Nat. Neurosci. 20, 864–871.  
1172 doi:10.1038/nn.4546
- 1173 Wyart, V., Tallon-Baudry, C., 2009. How Ongoing Fluctuations in Human Visual Cortex  
1174 Predict Perceptual Awareness: Baseline Shift versus Decision Bias. J. Neurosci. 29,  
1175 8715–8725. doi:10.1523/JNEUROSCI.0962-09.2009
- 1176

## Figure Legends

1177

1178 **Figure 1. General experimental design and behavioral performance in the working**  
1179 **memory task**

1180 **(A)** Experimental design. A subsequently masked target square was flashed in 1 out of 20  
1181 positions. Subjects were asked to report this location after a delay of up to 4s and to rate the  
1182 visibility of the target on a 4-point scale. A visible distractor square with features otherwise  
1183 identical to the target was shown on 50% of the trials during the retention period (at 1.75s). In  
1184 a perception-only control condition, the maintenance phase and location response were  
1185 omitted, and subjects assessed the visibility of the target immediately after the mask.

1186 **(B)** Spatial distributions of forced-choice localization performance in the working memory  
1187 task (experiment 1; 0 = correct target location; positive = clockwise offset). Error bars  
1188 indicate standard error of the mean (SEM) across subjects. The horizontal, dotted line  
1189 illustrates chance-level at 5%. Percentages show proportion of target-present trials from a  
1190 given visibility category. Due to low number of trials in individual visibility ratings 2, 3, and  
1191 4, all *seen* categories were collapsed for analyses.

1192

1193 **Figure 2. Behavioral evidence for non-conscious working memory**

1194 Spatial distributions of responses (0 = correct target location; positive = clockwise offset) as a  
1195 function of visibility and distractor presence **(A)**, conscious working memory load **(B)** and  
1196 delay duration **(C)**. Insets show rate of correct responding (within +/- 2 positions of actual  
1197 location) and precision of working memory representation separately for seen and unseen  
1198 trials. Error bars represent standard error of the mean (SEM) across subjects and horizontal,  
1199 dotted line indicates chance-level (5%). \* $p < .05$ , \*\* $p < .01$ , and \*\*\* $p < .001$  in a paired  
1200 sample  $t$ -test. Del = delay, Dis = distractor, L = load.

1201 The following figure supplement is available for Figure 2:

1202 Figure 2 – Supplement 1. Perceptual sensitivity does not correlate with working memory  
1203 performance on unseen trials.

1204

1205 **Figure 3. Neural signatures for conscious perception and maintenance in working**  
1206 **memory**

1207 **(A)** Sequence of brain activations (-200 – 800ms) evoked by consciously perceiving the target  
1208 in the perception (top) and working memory (bottom) task. Each topography depicts the  
1209 difference in amplitude between seen and unseen trials over a 100ms time window centered  
1210 on the time points shown (magnetometers only).

1211 **(B)** Average time courses of seen and unseen trials (-200 – 800ms) after subtraction of target-  
1212 absent trials in a group of parietal magnetometers in the perception (left) and working  
1213 memory (right) task. Shaded area illustrates standard error of the mean (SEM) across subjects.  
1214 Significant differences between conditions are depicted with a horizontal, black line  
1215 (Wilcoxon signed-rank test across subjects, uncorrected). For display purposes, data were  
1216 lowpass-filtered at 8Hz. T = target onset.

1217 **(C)** Temporal generalization matrices for decoding of visibility category as a function of  
1218 training and testing task. In each panel, a classifier was trained at every time sample (y-axis)  
1219 and tested on all other time points (x-axis). The diagonal gray line demarks classifiers trained  
1220 and tested on the same time sample. Please note the event markers in any panel involving the  
1221 perception task: Mean reaction time (target-present trials) for the visibility response is  
1222 indicated as vertical and/or horizontal, dotted lines. Any classifier beyond this point only  
1223 reflects post-visibility processes. Time courses of diagonal decoding and of classifiers  
1224 averaged over the P3b time window (0.3 – 0.6s) and over the working memory maintenance  
1225 period (0.8 – 2.5s) are shown as black, red, and blue insets. Thick lines indicate significant,  
1226 above-chance decoding of visibility (Wilcoxon signed-rank test across subjects, uncorrected,

1227 two-tailed except for diagonal). For display purposes, data were smoothed using a moving  
1228 average with a window of eight samples. AUC = area under the curve.

1229

1230 **Figure 4. A sustained decrease in alpha/beta power as a marker of conscious working**  
1231 **memory**

1232 **(A)** Average time-frequency power relative to baseline (dB) as a function of task and  
1233 visibility category in a group of occipital (left) and fronto-central (right) magnetometers.

1234 Mean reaction time (target-present trials) for the visibility response in the perception task is  
1235 indicated as a vertical, dotted line.

1236 **(B)** Beta band activity (13 – 30Hz; 0 – 2.1s) related to conscious working memory (seen –  
1237 unseen trials) as shown in magnetometers (top) and source space (bottom; in dB relative to  
1238 baseline). Black asterisks indicate sensors showing a significant difference as assessed by a  
1239 Monte-Carlo permutation test.

1240 **(C)** Same as in (A) and (B) but for unseen correct and unseen incorrect trials in the alpha band  
1241 (8 – 12Hz).

1242 The following figure supplement is available for Figure 4:

1243 Figure 4 – Supplement 1. Alpha- and beta-band desynchronizations serve as a general  
1244 signature of conscious processing and conscious working memory.

1245 Figure 4 – Supplement 2. Seen and unseen correct trials do not share the same discriminative  
1246 decoding axis.

1247 Figure 4 – Supplement 3. Bayesian statistics for the time-frequency analyses.

1248

1249 **Figure 5. Tracking the contents of conscious and non-conscious working memory**

1250 **(A)** Topographies (top) and time courses (bottom; -0.2 – 2.5s) of average circular-linear  
1251 correlations between the amplitude of the MEG signal (gradiometers) and target/distractor



1252 location. Shaded area demarks standard error of the mean (SEM) across subjects. Thick line  
1253 represents significant increase in correlation coefficient as compared to an empirical baseline  
1254 (one-tailed Wilcoxon signed-rank test across subjects, uncorrected).

1255 **(B)** Average time courses (-0.2 – 2.5s) of circular-linear correlation coefficients between  
1256 amplitude of the ERFs and target location as a function of visibility in the working memory  
1257 task in a group of left temporo-occipital (left), occipital (middle), and right temporo-occipital  
1258 (right) gradiometers. Shaded area demarks standard error of the mean (SEM) across subjects.  
1259 Thick line represents significant increase in correlation coefficient as compared to an  
1260 empirical baseline (one-tailed Wilcoxon signed-rank test across subjects, uncorrected). Insets  
1261 show average correlation coefficients (relative to an empirical baseline) over four time  
1262 windows: 0.1 – 0.3s (early), 0.3 – 0.6s (P3b), 0.6 – 1.55s (Del1), and 1.55 – 2.5s (Del2).  
1263 White asterisks denote significant differences to baseline (one-tailed Wilcoxon signed-rank  
1264 test across subjects), black asterisks significant differences between conditions (two-tailed  
1265 Wilcoxon signed-rank test across subjects). For display purposes, data were lowpass-filtered  
1266 at 8Hz. \* $p < .05$ , \*\* $p < .01$ , and \*\*\* $p < .001$ . Del1= first part of delay, Del2 = second part of  
1267 delay, T = target onset.

1268 **(C)** Same as in (B), but as a function of accuracy on the unseen trials (correct = within +/-2  
1269 positions of the target).

1270 The following figure supplement is available for Figure 5:

1271 Figure 5 – Supplement 1. Representation of seen target locations during conscious perception  
1272 and working memory.

1273 Figure 5 – Supplement 2. Circular-linear correlations and multivariate decoding reveal similar  
1274 time courses for target location.

1275 Figure 5 – Supplement 3. Tracking target/response location on unseen correct and incorrect  
1276 trials with multivariate decoding.

1277

1278 **Figure 6. Tracking response location in conscious and non-conscious working memory**

1279 (A) Topographies of average circular-linear correlations between the amplitude of the MEG

1280 signal (gradiometers) and response location. R = onset of the response screen.

1281 (B) Average time courses (left: stimulus-locked, -0.2 – 2.5s; right: response-locked, -0.5 –

1282 0.8s) of circular-linear correlation coefficients between amplitude of the ERFs and response

1283 location as a function of visibility in the working memory task in a group of occipital (top,

1284 left), frontal (top, right) left temporo-occipital (bottom, left) and right temporo-occipital

1285 (bottom, right) gradiometers. Shaded area demarks standard error of the mean (SEM) across

1286 subjects. Thick line represents significant increase in correlation coefficient as compared to an

1287 empirical baseline (one-tailed Wilcoxon signed-rank test across subjects, uncorrected). Insets

1288 show average correlation coefficients (relative to an empirical baseline) over four stimulus-

1289 locked time windows, 0.1 – 0.3s (early), 0.3 – 0.6s (P3b), 0.6 – 1.55s (Del1), and 1.55 – 2.5s

1290 (Del2), and two response-locked time windows, -0.5 – 0.0s (Del3) and 0.0 – 0.8s (Resp).

1291 White asterisks denote significant differences to baseline (one-tailed Wilcoxon signed-rank

1292 test across subjects), black asterisks significant differences between conditions (two-tailed

1293 Wilcoxon signed-rank test across subjects). For display purposes, data were lowpass-filtered

1294 at 8Hz. \* $p < .05$ , \*\* $p < .01$ , and \*\*\* $p < .001$ . Del1= first part of delay, Del2 = second part of

1295 delay, Del3 = last 500ms before response screen, R = response screen onset, T = target onset.

1296 (C) Same as in (B), but as a function of accuracy on the unseen trials (correct = within +/-2

1297 positions of the target).

1298 The following figure supplement is available for Figure 6:

1299 Figure 6 – Supplement 1. Topographies for circular-linear correlations with response location

1300 as a function of visibility.

1301 Figure 6 – Supplement 2. Circular-linear correlations and multivariate decoding reveal similar  
1302 time courses for response location.

1303

1304 **Figure 7. Activity-silent neural mechanisms underlying conscious and non-conscious**  
1305 **working memory**

1306 **(A)** Structure of a one-dimensional continuous attractor neural network (CANN). Neuronal  
1307 connections  $J(\theta, \theta')$  are translation-invariant in the space of the neurons' preferred stimulus  
1308 values  $(-\pi, \pi)$ , allowing the network to hold a continuous family of stationary states (bumps).  
1309 An external input  $I_e(\theta, t)$  containing the stimulus information triggers a bump state (red  
1310 curve) at the corresponding location in the network.

1311 **(B)** Model of a synaptic connection with short-term potentiation. In response to a presynaptic  
1312 spike train (bottom), the neurotransmitter release probability  $u$  increases and the fraction of  
1313 available neurotransmitter  $x$  decreases (middle), representing synaptic facilitation and  
1314 depression. Effective synaptic efficacy is proportional to  $ux$  (top).

1315 **(C)** Firing rate of neurons (top) and sequence of events (bottom; target and mask signal) when  
1316 simulating conscious working memory with  $A_{mask} = 50\text{Hz} < A_{critical}$ .

1317 **(D)** Same as in (C) for non-conscious working memory when  $A_{mask} = 65\text{Hz} > A_{critical}$ .

1318 **(E, F)** Performance of the network (distribution of responses) when mask amplitude was near  
1319 the critical level,  $A_{mask} = 62\text{Hz} \sim A_{critical}$ , and noise had been added to the system. Out of 4000  
1320 trials, 2035 resulted in the conscious (E) and the remainder in the non-conscious regime (F).  
1321 In both cases, performance remained above chance with the responses concentrated around  
1322 the initial target location.

1323

1324 **Figure 2 – Figure Supplement 1. Perceptual sensitivity does not correlate with working**  
1325 **memory performance on unseen trials**

1326 (A) Scatter plots depicting the relationship between detection  $d'$  and accuracy (left), the rate  
1327 of correct responding (middle), and precision (right) in the working memory task of  
1328 experiment 1 as a function of visibility.

1329 (B) Same as in (A), but for experiment 2.

1330

1331 **Figure 4 – Figure Supplement 1. Alpha- and beta-band desynchronizations serve as a**  
1332 **general signature of conscious processing and conscious working memory**

1333 (A) Perception task: Topographies represent the power difference (magnetometers) for seen  
1334 vs target-absent trials (top), seen vs unseen trials (middle), and unseen vs target-absent trials  
1335 (bottom) in the alpha (8 – 12Hz) and beta (13 – 30Hz) frequency bands as a function of time  
1336 (0 – 2.1s). Black asterisks indicate sensors showing a significant difference as assessed by a  
1337 cluster-based permutation test.

1338 (B) Working memory task: Topographies and panels are as in (A).

1339 (C) Working memory task: Topographies represent the power difference (magnetometers) for  
1340 unseen correct vs target-absent trials (top), unseen incorrect vs target-absent trials (middle),  
1341 and unseen correct vs unseen incorrect trials (bottom) in the alpha (8 – 12Hz) and beta (13 –  
1342 30Hz) frequency bands as a function of time (0 – 2.1s). Black asterisks indicate sensors  
1343 showing a significant difference as assessed by a cluster-based permutation test.

1344

1345 **Figure 4 – Figure Supplement 2. Seen and unseen correct trials do not share the same**  
1346 **discriminative decoding axis**

1347 (A) Temporal generalization matrices for a decoder trained on the ERFs to distinguish seen  
1348 from unseen trials in the perception task and tested in the working memory task, either with  
1349 the same labels (visibility decoder; left) or the unseen correct and incorrect trials (accuracy  
1350 decoder; right). In each panel, a classifier was trained at every time sample (y-axis) and tested

1351 on all other time points (x-axis). The diagonal gray line demarks classifiers trained and tested  
1352 on the same time sample. Please note the additional event marker: Mean reaction time (target-  
1353 present trials) for the visibility response is indicated as a horizontal, dotted line. Any classifier  
1354 beyond this point only reflects post-visibility processes. Time courses of diagonal decoding  
1355 are shown as black insets. Thick lines indicate significant, above-chance decoding (Wilcoxon  
1356 signed-rank test across subjects, uncorrected, one-tailed). For display purposes, data were  
1357 smoothed using a moving average with a window of eight samples. AUC = area under the  
1358 curve.

1359 **(B)** Same as in (A), except that the decoder was trained and tested on average power (relative  
1360 to baseline) in the alpha band (8 – 12 Hz). For display purposes, data were smoothed using a  
1361 moving average with a window of one sample.

1362 **(C)** Same as in (B), except that the decoder was trained and tested on average power (relative  
1363 to baseline) in the beta band (13 – 30 Hz).

1364

### 1365 **Figure 4 – Figure Supplement 3. Bayesian statistics for the time-frequency analyses**

1366 **(A)** Time courses of average alpha band activity (8 – 12Hz; -0.2 – 2.1s) in a group of frontal  
1367 sensors as a function of visibility (left) and accuracy on the unseen trials (right; correct =  
1368 within +/- 2 positions of the actual target location). Shaded area demarks standard error of the  
1369 mean (SEM) across subjects. Insets show Bayes Factors (as assessed in a two-tailed *t*-test) in  
1370 four time windows: 0.1 – 0.3s (early), 0.3 – 0.6s (P3b), 0.6 – 1.55s (Del1), and 1.55 – 2.1s  
1371 (Del2). Del1 = first part of the delay, Del2 = second part of the delay, T = target onset.

1372 **(B)** Same as in (A), but for average beta band (13 – 30Hz) activity.

1373

### 1374 **Figure 5 – Figure Supplement 1. Representation of seen target locations during** 1375 **conscious perception and working memory**

1376 Average time courses of circular-linear correlation coefficients between amplitude of the  
1377 ERFs and target location on seen trials as a function of task (perception and working memory)  
1378 in a group of left temporo-occipital (left), occipital (middle), and right temporo-occipital  
1379 (right) gradiometers. Shaded area demarks standard error of the mean (SEM) across subjects.  
1380 Mean reaction time (target-present trials) for the visibility response in the perception task is  
1381 indicated as a vertical, dotted line. Thick line represents significant increase in correlation  
1382 coefficient as compared to an empirical baseline (one-tailed Wilcoxon signed-rank test across  
1383 subjects, uncorrected). Insets show average correlation coefficients (relative to baseline) over  
1384 four time windows: 0.1 – 0.3s (early), 0.3 – 0.6s (P3b), 0.6 – 1.55s (Del1), and 1.55 – 2.5s  
1385 (Del2). White asterisks denote significant differences to baseline (one-tailed Wilcoxon  
1386 signed-rank test across subjects), black asterisks significant differences between conditions  
1387 (two-tailed Wilcoxon signed-rank test across subjects). For display purposes, data were  
1388 lowpass-filtered at 8Hz. \* $p < .05$ , \*\* $p < .01$ , and \*\*\* $p < .001$ . Del1 = first part of the delay  
1389 period, Del2 = second part of the delay period, T = target onset.

1390

1391 **Figure 5 – Figure Supplement 2. Circular-linear correlations and multivariate decoding**  
1392 **reveal similar time courses for target location**

1393 **(A)** Average time courses of circular-linear correlation coefficients between amplitude of the  
1394 ERFs and target location as a function of task (perception and working memory) and visibility  
1395 (seen and unseen) in a group of left temporo-occipital gradiometers. Shaded area demarks  
1396 standard error of the mean (SEM) across subjects. Thick line represents significant increase in  
1397 correlation coefficient as compared to an empirical baseline (one-tailed Wilcoxon signed-rank  
1398 test across subjects, uncorrected). Insets show average correlation coefficients (relative to  
1399 baseline) over four time windows: 0.1 – 0.3s (early), 0.3 – 0.6s (P3b), 0.6 – 1.55s (Del1), and  
1400 1.55 – 2.5s (Del2). White asterisks denote significant differences to baseline (one-tailed

1401 Wilcoxon signed-rank test across subjects). For display purposes, data were lowpass-filtered  
1402 at 8Hz.  $*p < .05$ ,  $**p < .01$ , and  $***p < .001$ . Del1 = first part of the delay period, Del2 =  
1403 second part of the delay period, T = target onset.

1404 **(B)** Average time courses of a linear support vector regression trained to predict target angle  
1405 as a function of task (perception and working memory) and visibility (seen and unseen). Thick  
1406 line represents significant increase in decoding accuracy (in radians) as compared to a  
1407 baseline (one-tailed Wilcoxon signed-rank test across subjects, uncorrected). Insets show  
1408 average correlation coefficients (relative to baseline) over four time windows: 0.1 – 0.3s  
1409 (early), 0.3 – 0.6s (P3b), 0.6 – 1.55s (Del1), and 1.55 – 2.5s (Del2). White asterisks denote  
1410 significant differences to baseline (one-tailed Wilcoxon signed-rank test across subjects). For  
1411 display purposes, data were lowpass-filtered at 8Hz.  $*p < .05$ ,  $**p < .01$ , and  $***p < .001$ .  
1412 Del1 = first part of the delay period, Del2 = second part of the delay period, T = target onset.

1413

1414 **Figure 5 – Figure Supplement 3. Tracking target/response location on unseen correct**  
1415 **and incorrect trials with multivariate decoding**

1416 **(A)** Average time courses of a linear support vector regression trained on seen correct trials to  
1417 predict target angle on the unseen correct (top) and unseen incorrect (bottom) trials. Thick line  
1418 represents significant increase in decoding accuracy (in radians) as compared to a baseline  
1419 (one-tailed Wilcoxon signed-rank test across subjects, uncorrected). Insets show average  
1420 correlation coefficients (relative to baseline) over four time windows: 0.1 – 0.3s (early), 0.3 –  
1421 0.6s (P3b), 0.6 – 1.55s (Del1), and 1.55 – 2.5s (Del2). White asterisks denote significant  
1422 differences to baseline (one-tailed Wilcoxon signed-rank test across subjects). For display  
1423 purposes, data were lowpass-filtered at 8Hz.  $*p < .05$ ,  $**p < .01$ , and  $***p < .001$ . Del1 = first  
1424 part of the delay period, Del2 = second part of the delay period, T = target onset.

1425 **(B)** Same as in (A), but for response location.

1426

1427 **Figure 6 – Figure Supplement 1. Topographies for circular-linear correlations with**  
1428 **response location as a function of visibility**

1429 Topographies of circular-linear correlations with response location as a function of time for  
1430 seen (left) and unseen (right) trials. The first three time bins are relative to target, the last two  
1431 relative to response screen onset. R = response screen onset.

1432

1433 **Figure 6 – Figure Supplement 2. Circular-linear correlations and multivariate decoding**  
1434 **reveal similar time courses for response location**

1435 **(A)** Average time courses of circular-linear correlation coefficients between amplitude of the  
1436 ERFs and response location as a function of task (perception and working memory) and  
1437 visibility (seen and unseen) in a group of left temporo-occipital gradiometers. Shaded area  
1438 demarks standard error of the mean (SEM) across subjects. Thick line represents significant  
1439 increase in correlation coefficient as compared to an empirical baseline (one-tailed Wilcoxon  
1440 signed-rank test across subjects, uncorrected). Insets show average correlation coefficients  
1441 (relative to baseline) over four time windows: 0.1 – 0.3s (early), 0.3 – 0.6s (P3b), 0.6 – 1.55s  
1442 (Del1), and 1.55 – 2.5s (Del2). White asterisks denote significant differences to baseline (one-  
1443 tailed Wilcoxon signed-rank test across subjects). For display purposes, data were lowpass-  
1444 filtered at 8Hz. \* $p < .05$ , \*\* $p < .01$ , and \*\*\* $p < .001$ . Del1 = first part of the delay period,  
1445 Del2 = second part of the delay period, T = target onset.

1446 **(B)** Average time courses of a linear support vector regression trained to predict response  
1447 angle as a function of task (perception and working memory) and visibility (seen and unseen).  
1448 Thick line represents significant increase in decoding accuracy (in radians) as compared to a  
1449 baseline (one-tailed Wilcoxon signed-rank test across subjects, uncorrected). Insets show  
1450 average correlation coefficients (relative to baseline) over four time windows: 0.1 – 0.3s



1451 (early), 0.3 – 0.6s (P3b), 0.6 – 1.55s (Del1), and 1.55 – 2.5s (Del2). White asterisks denote  
1452 significant differences to baseline (one-tailed Wilcoxon signed-rank test across subjects). For  
1453 display purposes, data were lowpass-filtered at 8Hz. \* $p < .05$ , \*\* $p < .01$ , and \*\*\* $p < .001$ .  
1454 Del1 = first part of the delay period, Del2 = second part of the delay period, T = target onset.  
1455

1456  
1457

## Tables

| Training → Testing |             | 0.1 – 0.3s   |                         | 0.3 – 0.6s   |                         | 0.6 – 1.55s  |                         | 1.55 – 2.5s  |                   |
|--------------------|-------------|--------------|-------------------------|--------------|-------------------------|--------------|-------------------------|--------------|-------------------|
|                    |             | AUC (SEM)    | <i>p</i>                | AUC (SEM)    | <i>p</i>                | AUC (SEM)    | <i>p</i>                | AUC (SEM)    | <i>p</i>          |
| P → P              | Diagonal    | 0.53 (0.01)  | <b>.004<sup>a</sup></b> | 0.58 (0.01)  | <b>.001<sup>a</sup></b> | 0.56 (0.01)  | <b>.001<sup>a</sup></b> | 0.52 (0.01)  | .058 <sup>a</sup> |
|                    | P3b         | 0.51 (0.01)  | .152 <sup>b</sup>       | 0.55 (0.01)  | <b>.003<sup>b</sup></b> | 0.52 (0.01)  | .064 <sup>b</sup>       | 0.51 (0.005) | .101 <sup>b</sup> |
|                    | Maintenance | 0.50 (0.004) | .507 <sup>b</sup>       | 0.50 (0.005) | .382 <sup>b</sup>       | 0.52 (0.01)  | <b>.046<sup>b</sup></b> | 0.51 (0.01)  | .382 <sup>b</sup> |
| P → WM             | Diagonal    | 0.52 (0.005) | <b>.003<sup>a</sup></b> | 0.55 (0.01)  | <b>.001<sup>a</sup></b> | 0.53 (0.005) | <b>.001<sup>a</sup></b> | 0.50 (0.01)  | .486 <sup>a</sup> |
|                    | P3b         | 0.50 (0.004) | .279 <sup>b</sup>       | 0.53 (0.01)  | <b>.011<sup>b</sup></b> | 0.51 (0.01)  | .101 <sup>b</sup>       | 0.49 (0.01)  | .101 <sup>b</sup> |
|                    | Maintenance | 0.49 (0.005) | <b>.039<sup>b</sup></b> | 0.49 (0.01)  | .311 <sup>b</sup>       | 0.51 (0.006) | .279 <sup>b</sup>       | 0.50 (0.01)  | .972 <sup>b</sup> |
| WM → WM            | Diagonal    | 0.52 (0.01)  | .066 <sup>a</sup>       | 0.57 (0.02)  | <b>.007<sup>a</sup></b> | 0.55 (0.01)  | <b>.001<sup>a</sup></b> | 0.52 (0.01)  | .173 <sup>a</sup> |
|                    | P3b         | 0.50 (0.01)  | .807 <sup>b</sup>       | 0.54 (0.01)  | <b>.020<sup>b</sup></b> | 0.50 (0.01)  | .807 <sup>b</sup>       | 0.49 (0.01)  | .422 <sup>b</sup> |
|                    | Maintenance | 0.50 (0.01)  | .507 <sup>b</sup>       | 0.49 (0.01)  | .552 <sup>b</sup>       | 0.51 (0.01)  | .650 <sup>b</sup>       | 0.51 (0.01)  | .600 <sup>b</sup> |
| WM → P             | Diagonal    | 0.52 (0.005) | <b>.010<sup>a</sup></b> | 0.55 (0.01)  | <b>.001<sup>a</sup></b> | 0.53 (0.01)  | <b>.014<sup>a</sup></b> | 0.51 (0.01)  | .276 <sup>a</sup> |
|                    | P3b         | 0.50 (0.006) | .753 <sup>b</sup>       | 0.53 (0.01)  | <b>.016<sup>b</sup></b> | 0.50 (0.01)  | .972 <sup>b</sup>       | 0.49 (0.01)  | .463 <sup>b</sup> |
|                    | Maintenance | 0.49 (0.01)  | .279 <sup>b</sup>       | 0.49 (0.01)  | .101 <sup>b</sup>       | 0.51 (0.01)  | .701 <sup>b</sup>       | 0.50 (0.01)  | .972 <sup>b</sup> |

1458

1459 **Table 1. Statistics for decoding analyses**

1460

1461 Statistics are shown for decoding of visibility category (seen vs. unseen) as a function of task

1462 and testing time bin. The first column identifies the respective training and testing sets (P =

1463 perception task; WM = working memory task), the second column the training classifiers

1464 (Diagonal = diagonal, P3b = 300 – 600ms, Maintenance = 0.8 – 2.5s), that were averaged.

1465 Bold numbers indicate above-chance decoding performance (<sup>a</sup>one-tailed, <sup>b</sup>two-tailed

1466 Wilcoxon signed-rank test across subjects). AUC = area under the curve; SEM = standard

1467 error of the mean (across participants).

1468

1469

|            |         | 0.1 – 0.3s: Target |                | 0.3 – 0.6s: Target |                               | 0.6 – 1.55s: Target |                               | 1.55 – 2.5s: Target |                               | -0.5 – 0s: Resp    |                              | 0 – 0.8s: Resp     |               |       |
|------------|---------|--------------------|----------------|--------------------|-------------------------------|---------------------|-------------------------------|---------------------|-------------------------------|--------------------|------------------------------|--------------------|---------------|-------|
|            |         | $\Delta\rho$ (SEM) | <i>p</i>       | $\Delta\rho$ (SEM) | <i>p</i>                      | $\Delta\rho$ (SEM)  | <i>p</i>                      | $\Delta\rho$ (SEM)  | <i>p</i>                      | $\Delta\rho$ (SEM) | <i>p</i>                     | $\Delta\rho$ (SEM) | <i>p</i>      |       |
| Distractor | All     | Occ                | 0.021 (0.005)  | <.001              | 0.015 (0.005)                 | .009                | n/a                           | n/a                 | n/a                           | n/a                | n/a                          | n/a                | n/a           |       |
| Target     | All     | Occ                | 0.021 (0.005)  | <.001              | 0.030 (0.007)                 | <.001               | 0.004 (0.002)                 | .064                | 0.002 (0.020)                 | .108               | n/a                          | n/a                | n/a           |       |
|            | P Seen  | L Temp             | 0.011 (0.004)  | .007               | 0.014 (0.005)                 | .009                | 0.005 (0.002)                 | .029                | -0.002 (0.002)                | .830               | n/a                          | n/a                | n/a           |       |
|            |         | Occ                | 0.004 (0.004)  | .207               | 0.019 (0.003)                 | <.001               | 0.003 (0.003)                 | .153                | -0.001 (0.002)                | .751               | n/a                          | n/a                | n/a           |       |
|            |         | R Temp             | 0.012 (0.004)  | .011               | 0.030 (0.004)                 | <.001               | 0.008 (0.002)                 | .001                | -0.003 (0.002)                | .905               | n/a                          | n/a                | n/a           |       |
|            | Seen    | L Temp             | 0.018 (0.005)  | <.001              | 0.021 (0.004)                 | <.001               | 0.006 (0.003)                 | .024                | -0.002 (0.002)                | .729               | n/a                          | n/a                | n/a           |       |
|            |         | Occ                | 0.024 (0.006)  | .001               | 0.031 (0.008)                 | <.001               | 0.009 (0.004)                 | .024                | 0.001 (0.003)                 | .446               | n/a                          | n/a                | n/a           |       |
|            |         | R Temp             | 0.016 (0.008)  | .064               | 0.031 (0.006)                 | <.001               | 0.005 (0.004)                 | .064                | 0.001 (0.003)                 | .342               | n/a                          | n/a                | n/a           |       |
|            | Unseen  | L Temp             | 0.009 (0.004)  | .047               | 0.007 (0.004)                 | .084                | -0.004 (0.002)                | .971                | -2.9*10 <sup>-4</sup> (0.002) | .527               | n/a                          | n/a                | n/a           |       |
|            |         | Occ                | 0.011 (0.005)  | .024               | 0.007 (0.004)                 | .055                | -0.006 (0.003)                | .966                | 3.0*10 <sup>-4</sup> (0.003)  | .446               | n/a                          | n/a                | n/a           |       |
|            |         | R Temp             | -0.002 (0.005) | .632               | 0.002 (0.002)                 | .527                | -0.003 (0.002)                | .936                | 0.002 (0.004)                 | .473               | n/a                          | n/a                | n/a           |       |
|            | Unseen+ | L Temp             | 0.014 (0.006)  | .040               | -0.002 (0.007)                | .682                | 0.002 (0.004)                 | .271                | -0.002 (0.005)                | .812               | n/a                          | n/a                | n/a           |       |
|            |         | Occ                | 0.007 (0.007)  | .227               | 0.007 (0.007)                 | .170                | -0.001 (0.006)                | .554                | 2.2*10 <sup>-4</sup> (0.005)  | .580               | n/a                          | n/a                | n/a           |       |
|            |         | R Temp             | -0.012 (0.008) | .803               | -0.010 (0.007)                | .905                | -0.004 (0.005)                | .706                | -0.005 (0.006)                | .812               | n/a                          | n/a                | n/a           |       |
|            | Unseen- | L Temp             | 0.006 (0.009)  | .500               | 0.002 (0.010)                 | .554                | -0.006 (0.004)                | .916                | 0.001 (0.005)                 | .393               | n/a                          | n/a                | n/a           |       |
|            |         | Occ                | 0.001 (0.005)  | .580               | -0.012 (0.006)                | .966                | -0.009 (0.004)                | .980                | -0.007 (0.003)                | .989               | n/a                          | n/a                | n/a           |       |
|            |         | R Temp             | -0.012 (0.007) | .971               | -0.003 (0.004)                | .847                | -0.006 (0.005)                | .916                | -8.6*10 <sup>-5</sup> (0.004) | .420               | n/a                          | n/a                | n/a           |       |
| Response   | Seen    | L Temp             | 0.015 (0.005)  | .002               | 0.020 (0.005)                 | .005                | 0.005 (0.002)                 | .020                | -1.2*10 <sup>-4</sup> (0.002) | .420               | -0.001 (0.003)               | .527               | 0.022 (0.005) | .001  |
|            |         | Occ                | 0.018 (0.007)  | .020               | 0.029 (0.007)                 | <.001               | 0.008 (0.003)                 | .007                | 0.003 (0.004)                 | .207               | 4.3*10 <sup>-4</sup> (0.003) | .446               | 0.020 (0.004) | <.001 |
|            |         | R Temp             | 0.014 (0.009)  | .122               | 0.030 (0.006)                 | <.001               | 0.005 (0.004)                 | .137                | 6.2*10 <sup>-4</sup> (0.003)  | .473               | 4.9*10 <sup>-4</sup> (0.005) | .318               | 0.025 (0.004) | <.001 |
|            | Frontal | L Temp             | 0.006 (0.005)  | .170               | 0.006 (0.004)                 | .122                | 0.006 (0.003)                 | .034                | 0.003 (0.003)                 | .073               | 0.004 (0.005)                | .249               | 0.034 (0.007) | <.001 |
|            |         | L Temp             | 0.006 (0.004)  | .084               | -0.001 (0.004)                | .500                | 0.003 (0.002)                 | .064                | 0.005 (0.003)                 | .137               | 0.004 (0.006)                | .294               | 0.012 (0.003) | <.001 |
|            |         | Occ                | 0.003 (0.005)  | .294               | -5.1*10 <sup>-4</sup> (0.003) | .394                | 0.003 (0.002)                 | .108                | 0.006 (0.004)                 | .096               | 0.009 (0.006)                | .170               | 0.017 (0.004) | <.001 |
|            |         | R Temp             | -0.003 (0.005) | .773               | -0.002 (0.006)                | .368                | -9.9*10 <sup>-4</sup> (0.003) | .446                | 9.7*10 <sup>-4</sup> (0.003)  | .393               | 0.002 (0.006)                | .473               | 0.015 (0.005) | .004  |
|            | Frontal | L Temp             | 0.001 (0.004)  | .473               | 0.009 (0.003)                 | .006                | 0.002 (0.003)                 | .137                | 0.007 (0.002)                 | .003               | 6.4*10 <sup>-4</sup> (0.004) | .682               | 0.027 (0.007) | .002  |
|            |         | L Temp             | 0.008 (0.007)  | .096               | 0.002 (0.007)                 | .446                | 0.005 (0.003)                 | .108                | 0.002 (0.005)                 | .446               | 0.014 (0.007)                | .055               | 0.012 (0.007) | .024  |
|            |         | Occ                | -0.003 (0.009) | .580               | 0.004 (0.008)                 | .393                | 0.002 (0.005)                 | .393                | 0.002 (0.005)                 | .473               | 0.014 (0.007)                | .170               | 0.006 (0.006) | .227  |
|            |         | R Temp             | -0.012 (0.008) | .878               | -0.005 (0.006)                | .773                | -0.004 (0.005)                | .729                | -0.005 (0.005)                | .830               | 0.008 (0.007)                | .170               | 0.014 (0.008) | .084  |
|            | Frontal | L Temp             | 0.001 (0.008)  | .318               | 0.006 (0.006)                 | .122                | -0.004 (0.004)                | .892                | 0.005 (0.004)                 | .342               | 0.005 (0.006)                | .227               | 0.024 (0.007) | .004  |
|            |         | L Temp             | 0.008 (0.005)  | .096               | 8.3*10 <sup>-4</sup> (0.005)  | .446                | 0.004 (0.003)                 | .153                | -0.003 (0.004)                | .207               | -0.003 (0.006)               | .632               | 0.011 (0.004) | .024  |
|            |         | Occ                | 0.013 (0.009)  | .122               | 0.009 (0.007)                 | .096                | 0.003 (0.003)                 | .153                | 0.003 (0.006)                 | .342               | 0.003 (0.005)                | .227               | 0.014 (0.006) | .020  |
|            |         | R Temp             | -0.005 (0.008) | .812               | -0.009 (0.005)                | .916                | -0.004 (0.004)                | .905                | -0.002 (0.005)                | .682               | 0.002 (0.005)                | .368               | 0.007 (0.006) | .137  |
|            | Frontal | L Temp             | -0.007 (0.006) | .863               | -0.003 (0.004)                | .729                | 0.004 (0.004)                 | .207                | -0.010 (0.003)                | .527               | -0.005 (0.003)               | .927               | 0.018 (.007)  | .011  |

1470  
1471  
1472  
1473  
1474  
1475  
1476  
1477  
1478  
1479  
1480  
1481  
1482  
1483

**Table 2. Statistics for circular-linear correlation analyses**

Statistics for circular-linear correlation analyses between the average amplitude of the MEG signal in the gradiometers and distractor, target, and response position are listed as a function of visibility, accuracy, channel group and time window. The first four time windows are relative to target onset, the last two relative to the onset of the response screen. Bold numbers indicate significant differences in correlation values relative to an empirical baseline (one-tailed Wilcoxon signed-rank test). Frontal = frontal gradiometers; L Temp = left temporo-occipital gradiometers; Occ = occipital gradiometers; P = perception task; Resp = response; R Temp = right temporo-occipital gradiometers; SEM = standard error of the mean (across subjects); Unseen+ = unseen correct trials (within +/- 2 positions of actual target location); Unseen- = unseen incorrect trials.

| BF         |         | 0.1 – 0.3s: Target | 0.3 – 0.6s: Target | 0.6 – 1.55s: Target | 1.55 – 2.5s: Target | -0.5 – 0: Resp | 0 – 0.8: Resp |
|------------|---------|--------------------|--------------------|---------------------|---------------------|----------------|---------------|
| Distractor |         |                    |                    |                     |                     |                |               |
| All        | Occ     | <b>46.72</b>       | <b>14.47</b>       | n/a                 | n/a                 | n/a            | n/a           |
| Target     |         |                    |                    |                     |                     |                |               |
| All        | Occ     | <b>109.60</b>      | <b>125.33</b>      | 2.29                | 0.77                | n/a            | n/a           |
| P Seen     | L Temp  | <b>4.84</b>        | <b>10.03</b>       | <b>3.62</b>         | 0.16                | n/a            | n/a           |
|            | Occ     | 0.68               | <b>496.64</b>      | 0.96                | 0.21                | n/a            | n/a           |
| Seen       | R Temp  | <b>6.81</b>        | <b>5256.15</b>     | <b>23.35</b>        | 0.13                | n/a            | n/a           |
|            | L Temp  | <b>24.07</b>       | <b>175.81</b>      | <b>3.89</b>         | 0.17                | n/a            | n/a           |
| Unseen     | Occ     | <b>56.90</b>       | <b>45.43</b>       | <b>3.77</b>         | 0.40                | n/a            | n/a           |
|            | R Temp  | 2.31               | <b>496.24</b>      | 1.22                | 0.38                | n/a            | n/a           |
|            | L Temp  | 2.58               | 1.33               | 0.11                | 0.28                | n/a            | n/a           |
| Unseen+    | Occ     | <b>3.34</b>        | 1.53               | 0.11                | 0.30                | n/a            | n/a           |
|            | R Temp  | 0.23               | 0.48               | 0.14                | 0.47                | n/a            | n/a           |
|            | L Temp  | <b>3.75</b>        | 0.22               | 0.46                | 0.21                | n/a            | n/a           |
| Unseen-    | Occ     | 0.76               | 0.65               | 0.24                | 0.29                | n/a            | n/a           |
|            | R Temp  | 0.13               | 0.13               | 0.18                | 0.16                | n/a            | n/a           |
|            | L Temp  | 0.49               | 0.32               | 0.13                | 0.34                | n/a            | n/a           |
|            | Occ     | 0.35               | 0.11               | 0.11                | 0.11                | n/a            | n/a           |
|            | R Temp  | 0.12               | 0.19               | 0.14                | 0.28                | n/a            | n/a           |
| Response   |         |                    |                    |                     |                     |                |               |
| Seen       | L Temp  | <b>11.35</b>       | <b>41.95</b>       | <b>4.16</b>         | 0.27                | 0.23           | <b>90.13</b>  |
|            | Occ     | <b>7.41</b>        | <b>56.46</b>       | <b>4.60</b>         | 0.56                | 0.31           | <b>152.61</b> |
|            | R Temp  | 1.42               | <b>330.43</b>      | 0.95                | 0.32                | 0.30           | <b>606.35</b> |
| Unseen     | Frontal | 1.05               | 1.12               | 2.42                | 0.57                | 0.60           | <b>127.23</b> |
|            | L Temp  | 1.28               | 0.23               | 1.37                | 1.05                | 0.52           | <b>137.36</b> |
|            | Occ     | 0.44               | 0.25               | 0.98                | 1.32                | 1.24           | <b>49.11</b>  |
| Unseen+    | R Temp  | 0.18               | 0.22               | 0.23                | 0.35                | 0.35           | <b>13.73</b>  |
|            | Frontal | 0.35               | <b>12.69</b>       | 0.47                | <b>15.86</b>        | 0.32           | <b>41.58</b>  |
|            | L Temp  | 0.85               | 0.33               | 2.19                | 0.38                | 2.25           | 2.05          |
| Unseen-    | Occ     | 0.23               | 0.43               | 0.38                | 0.35                | 2.10           | 0.66          |
|            | R Temp  | 0.13               | 0.17               | 0.18                | 0.17                | 0.78           | 1.90          |
|            | Frontal | 0.32               | 0.82               | 0.14                | 0.77                | 0.59           | <b>16.76</b>  |
| Unseen-    | L Temp  | 1.47               | 0.32               | 1.15                | 0.50                | 0.20           | <b>5.13</b>   |
|            | Occ     | 1.23               | 0.94               | 0.84                | 0.44                | 0.47           | <b>3.45</b>   |
|            | R Temp  | 0.19               | 0.12               | 0.16                | 0.21                | 0.36           | 0.88          |
|            | Frontal | 0.15               | 0.18               | 0.69                | 0.23                | 0.13           | <b>5.14</b>   |

1484

1485 **Table 3. Bayes Factors for circular-linear correlation analyses**

1486

1487 Bayes Factors for circular-linear correlation analyses between the average amplitude of the

1488 MEG signal in the gradiometers and distractor, target, and response position are shown as a

1489 function of visibility, accuracy, channel group and time window. The first four time windows

1490 are relative to target onset, the last two relative to the onset of the response screen. Bold

1491 numbers indicate strong evidence in favor of the alternative hypothesis (i.e., an increase in

1492 correlation values relative to an empirical baseline as assessed by a one-tailed Bayesian *t*-

1493 test). Frontal = frontal gradiometers; L Temp = left temporo-occipital gradiometers; Occ =

1494 occipital gradiometers; P = perception task; Resp = response; R Temp = right temporo-

1495 occipital gradiometers; SEM = standard error of the mean (across subjects); Unseen+ =

1496 unseen correct trials (within +/- 2 positions of actual target location); Unseen- = unseen

1497 incorrect trials.

| Task           | Target  | Visibility 4<br>(clearly seen)<br><i>M (SD)</i> | Visibility 3<br>(weakly seen)<br><i>M (SD)</i> | Visibility 2<br>(glimpse)<br><i>M (SD)</i> | Visibility 1<br>(unseen)<br><i>M (SD)</i> |
|----------------|---------|---|--|--|---|
| Perception     | Present | 9.8 (16.6)                                      | 17.3 (12.8)                                    | 46 (17.8)                                  | 61.2 (15.4)                               |
|                | Absent  | 0.1 (0.3)                                       | 0.5 (1.4)                                      | 3.0 (3.8)                                  | 28.9 (5.7)                                |
| Working memory | Present | 10.1 (18.0)                                     | 15.8 (14.6)                                    | 45.2 (17.8)                                | 57.4 (17.4)                               |
|                | Absent  | 0.0 (0.0)                                       | 0.2 (0.6)                                      | 1.9 (1.9)                                  | 29.3 (4.4)                                |

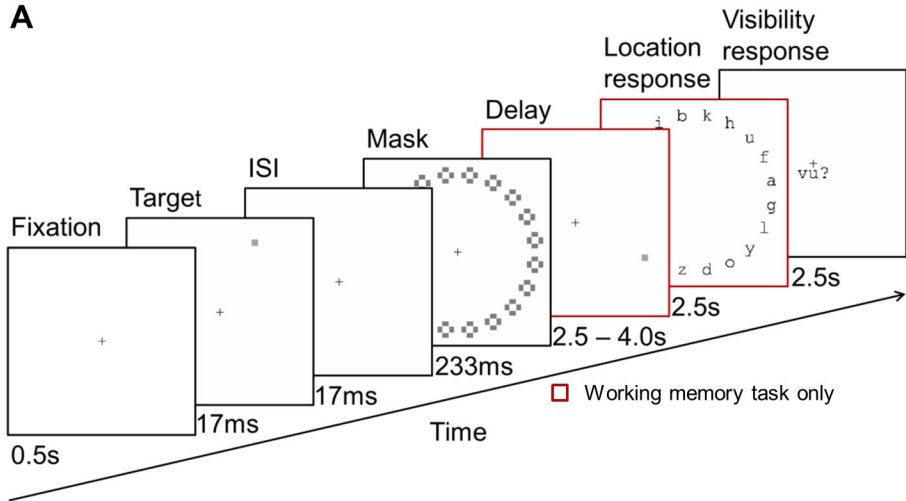
1498

1499 **Table 4. Trial counts**

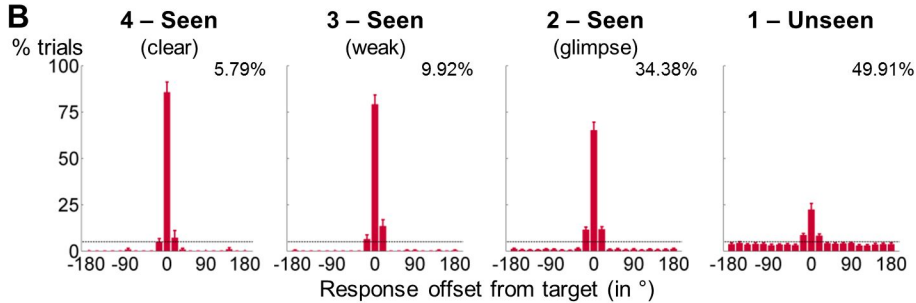
1500 Number of trials included in the MEG analyses are listed as a function of task (perception vs.  
1501 working memory task), target (present vs. absent), and visibility rating. Mean (*M*) and  
1502 standard deviation (*SD*) are based on 13 participants and all trials retained after preprocessing  
1503 of the MEG data.

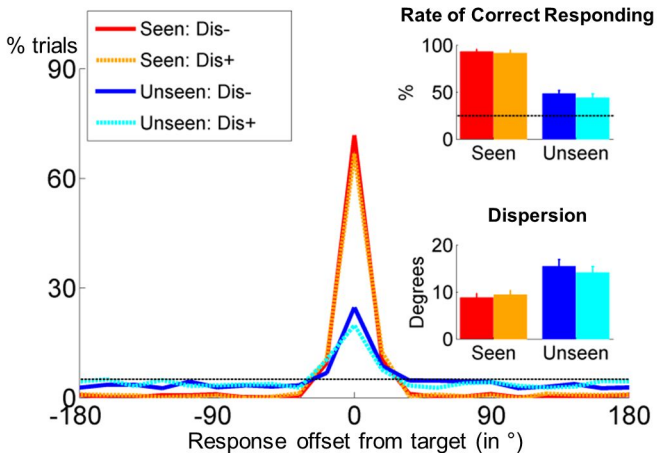
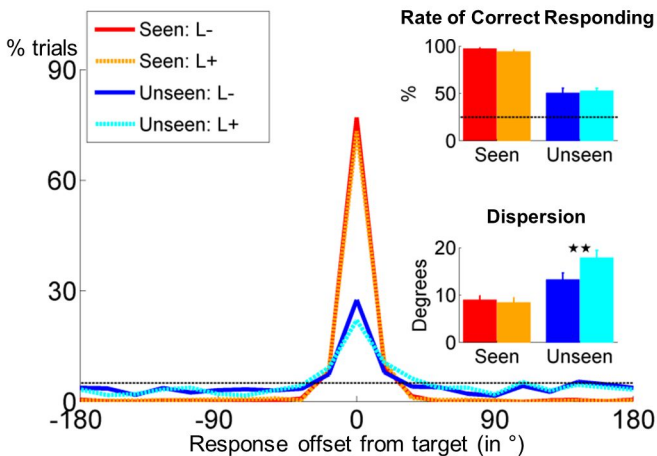
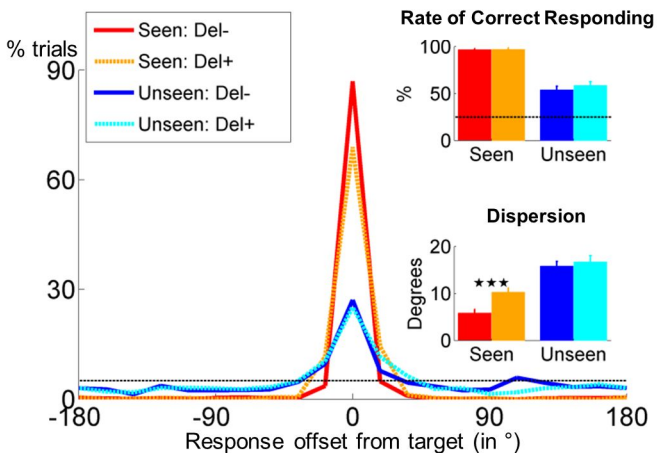
# Figure 1

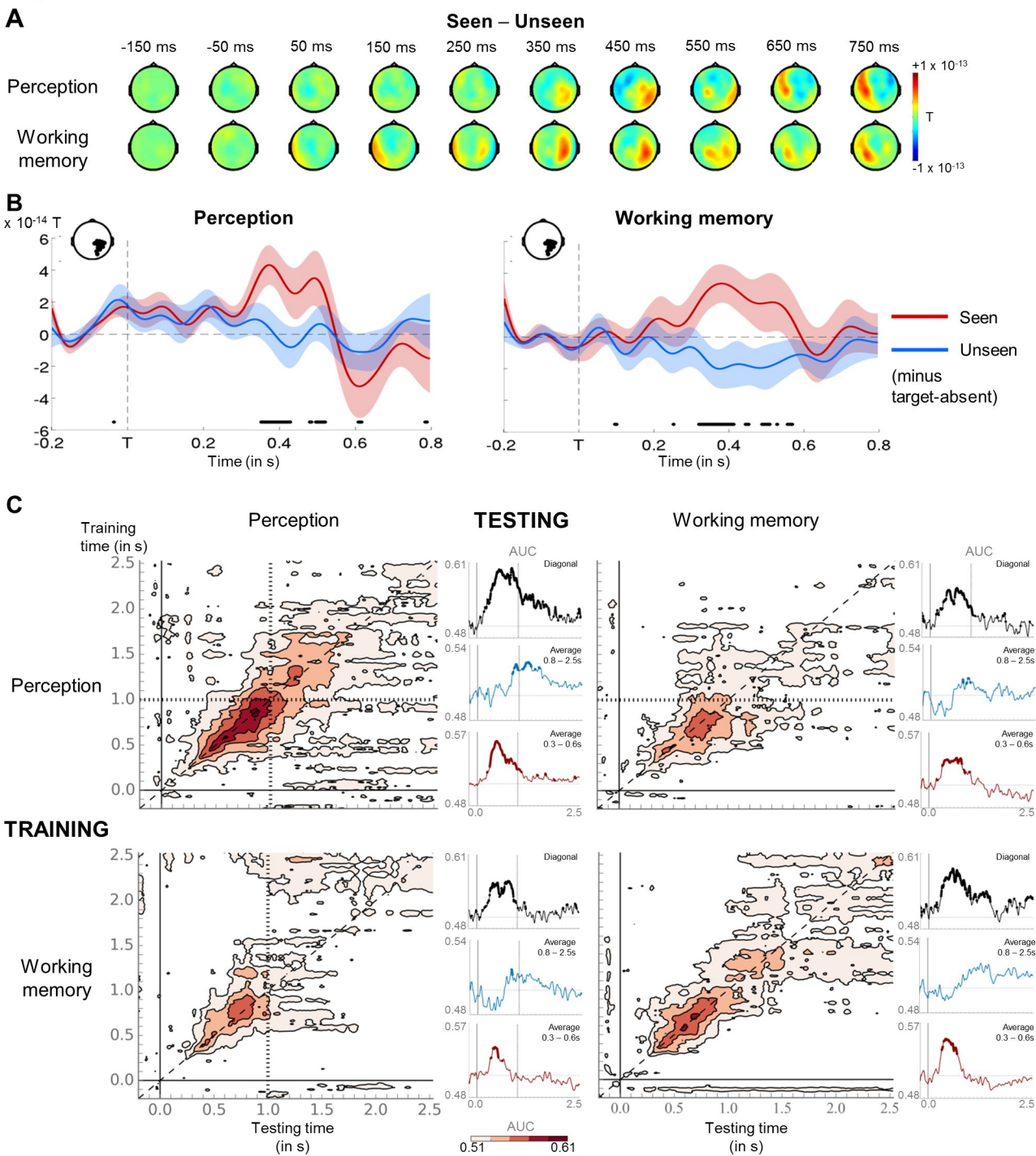
## A



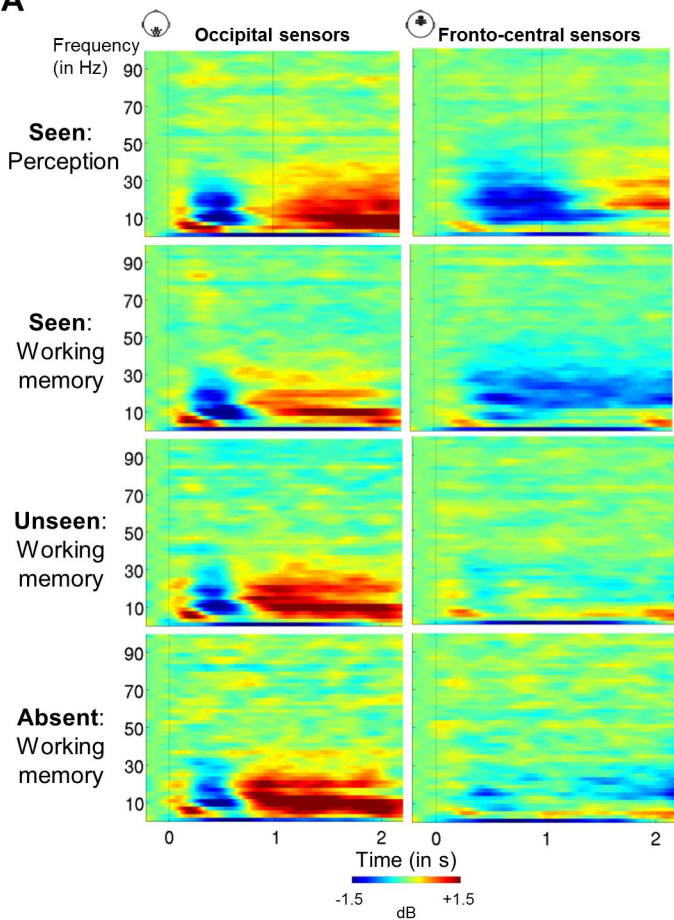
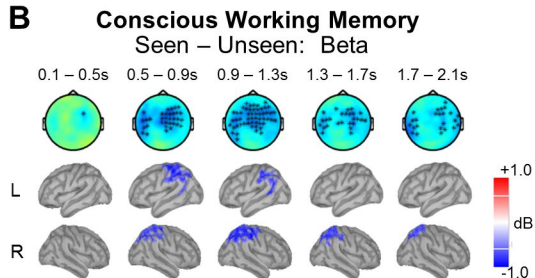
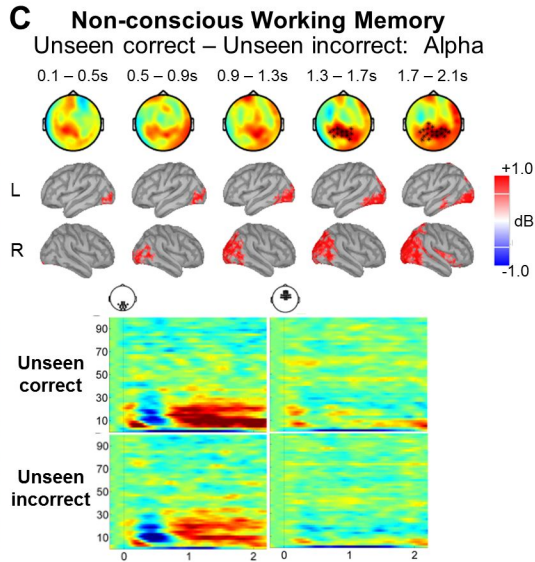
## B

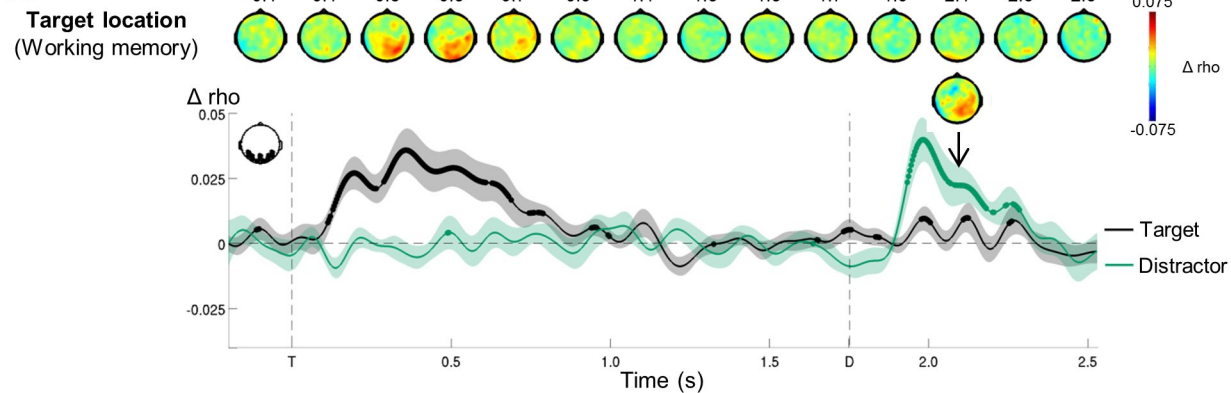
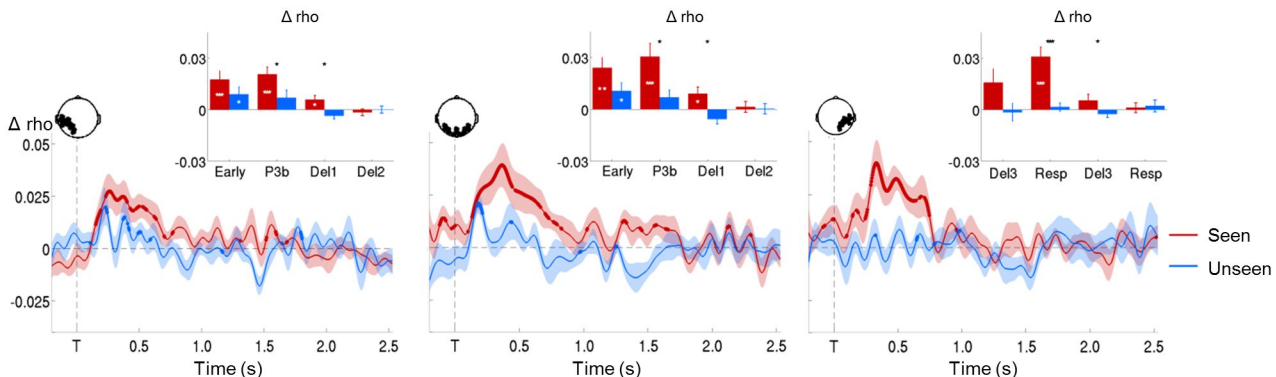
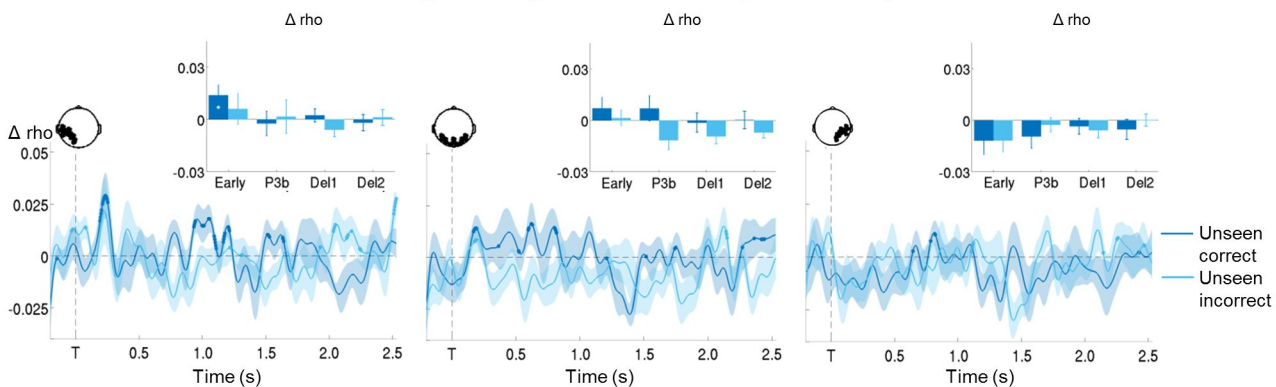


**Figure 2****A Effect of distractor****B Effect of working memory load****C Effect of working memory delay**

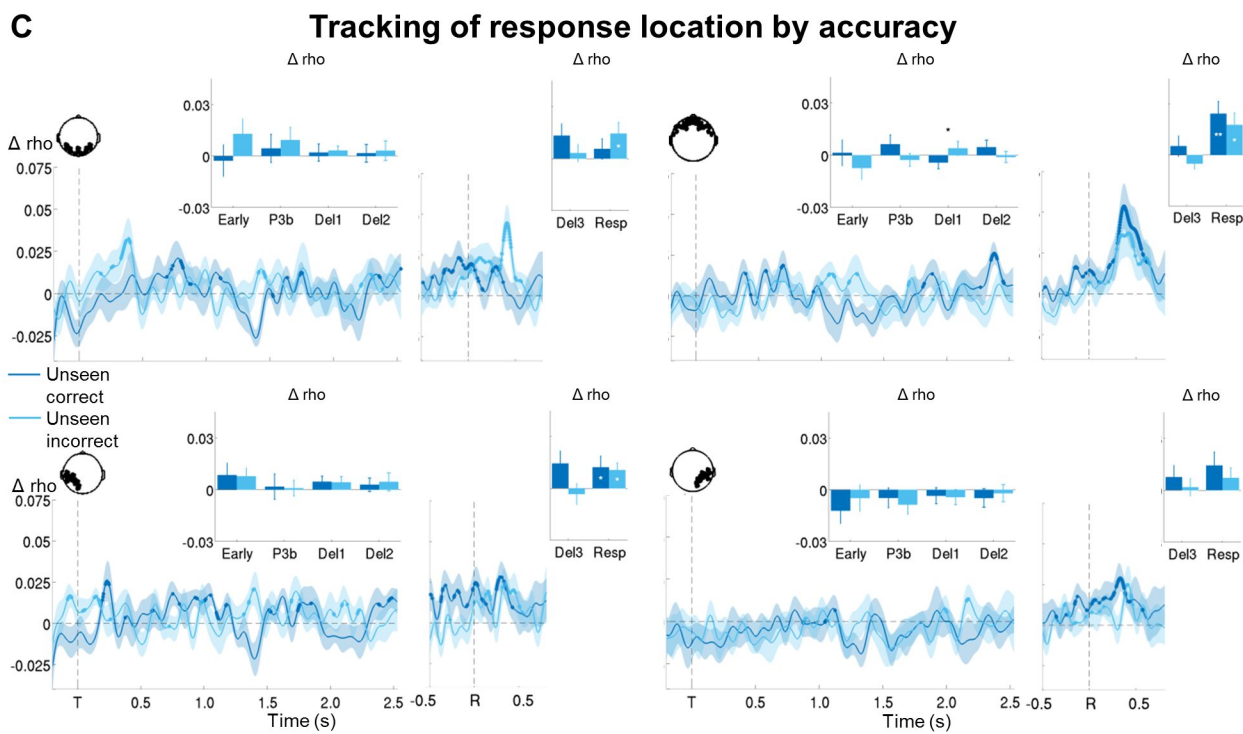
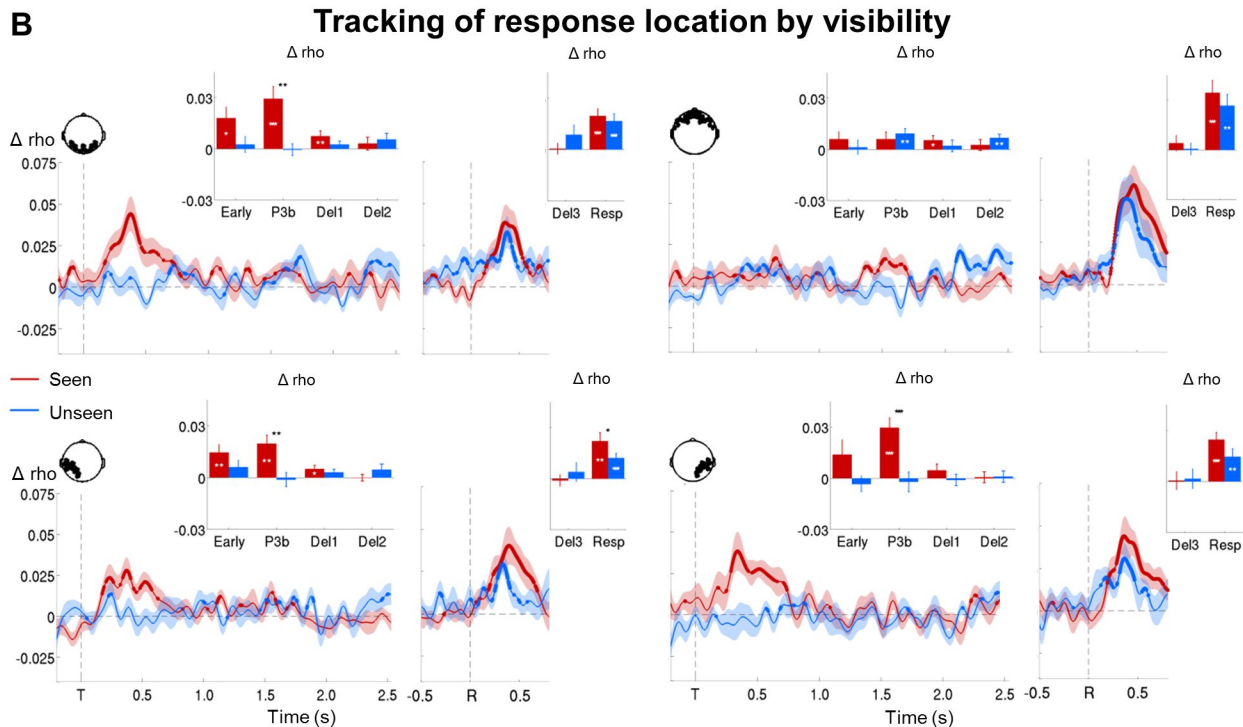
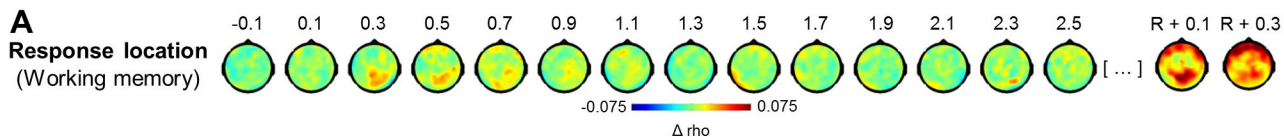
**Figure 3**

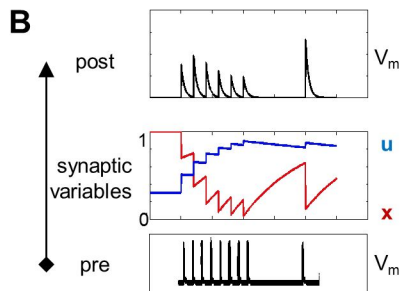
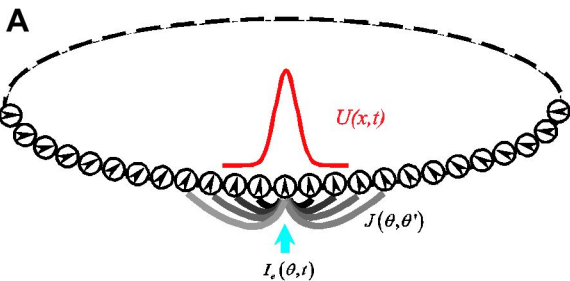
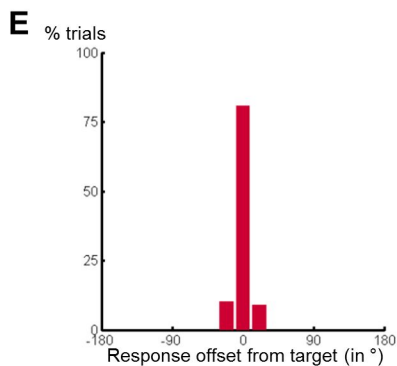
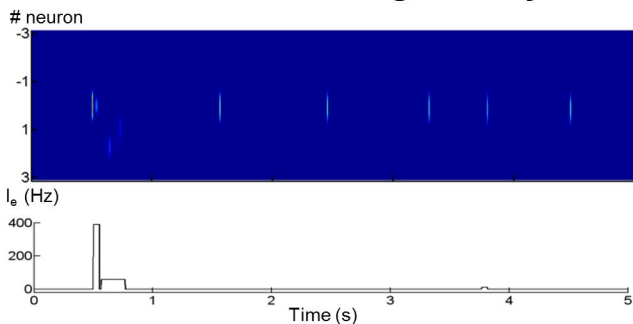
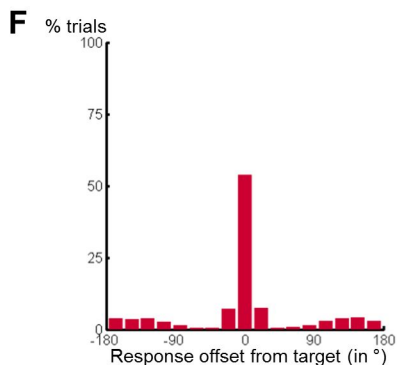
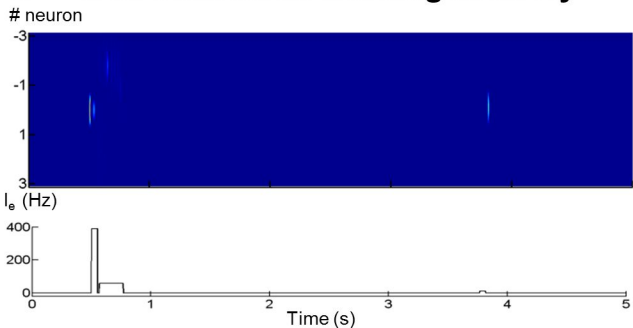


**Figure 4****A****B****C**

**Figure 5****A****B****Tracking of target location by visibility****C****Tracking of target location by accuracy**

# Figure 6

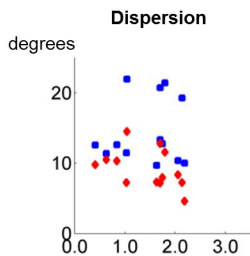
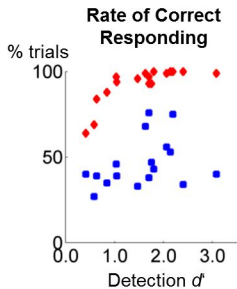
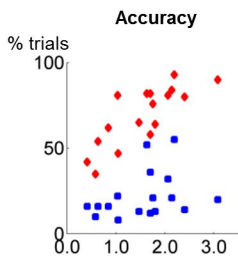


**Figure 7****C Conscious working memory****D Non-conscious working memory**

# Figure 2 – Figure Supplement 1

## A

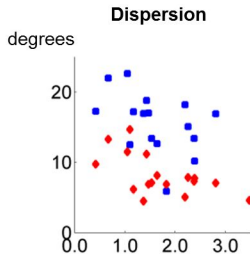
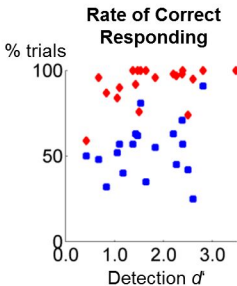
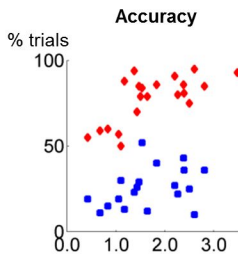
### Experiment 1



◆ Seen  
■ Unseen

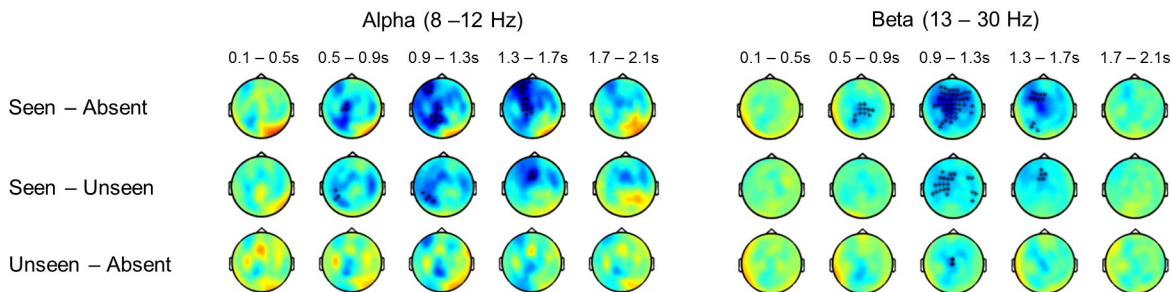
## B

### Experiment 2

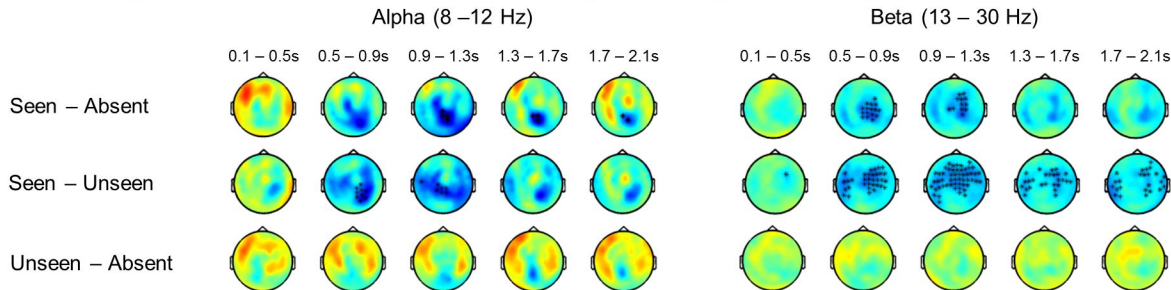


## Figure 4 – Figure Supplement 1

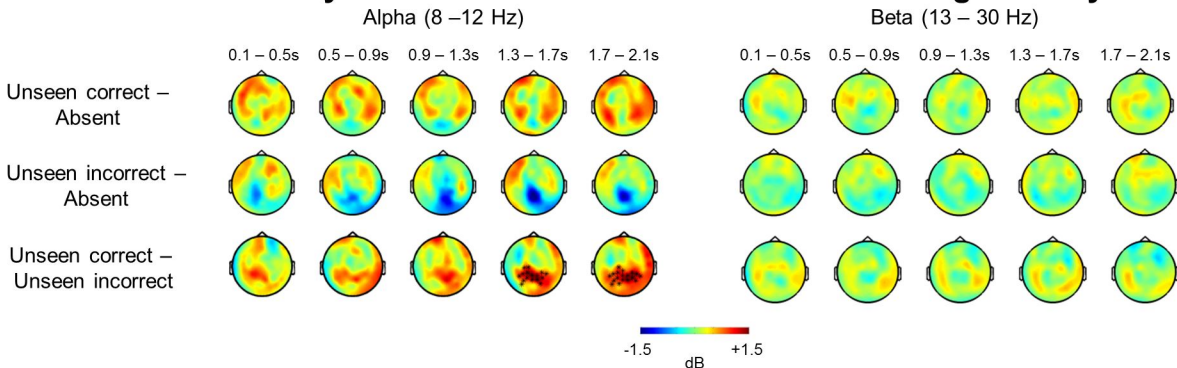
### A Signatures of conscious perception in the alpha/beta band



### B Signatures of conscious working memory in the alpha/beta band



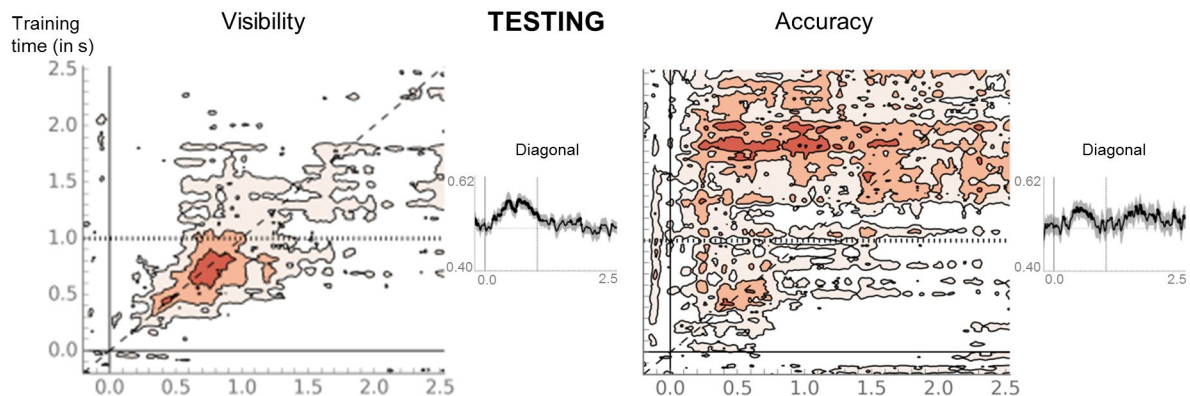
### C Absence of desynchronization for non-conscious working memory



# Figure 4 – Figure Supplement 2

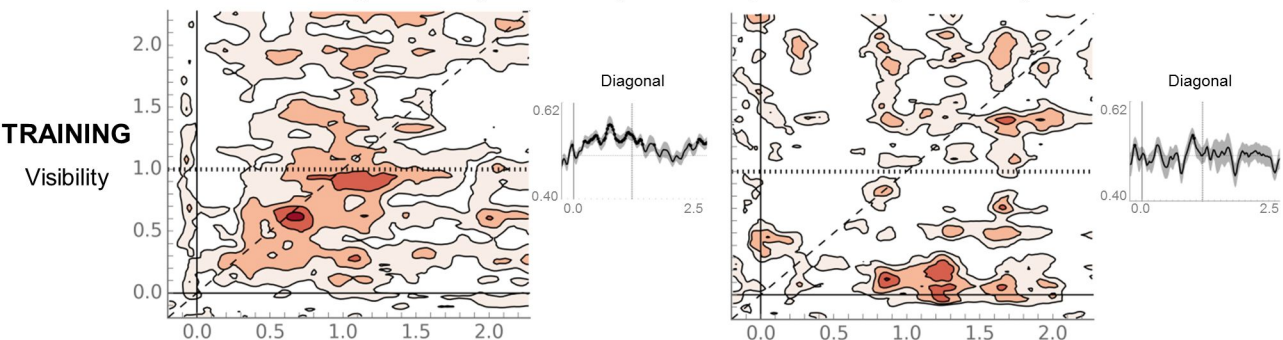
**A**

## Decoding visibility/accuracy in the ERF



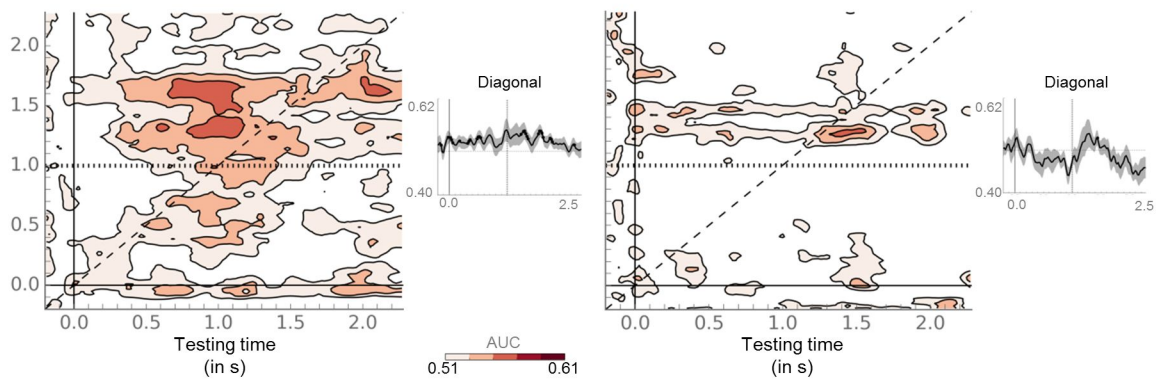
**B**

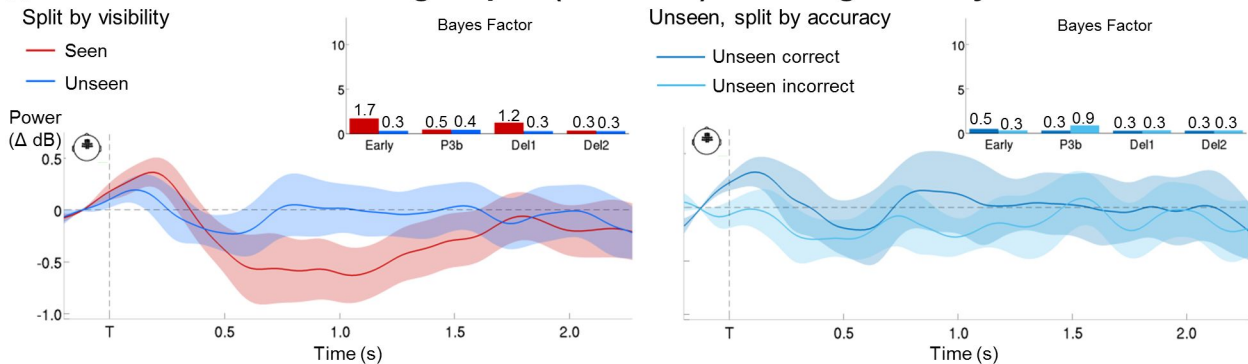
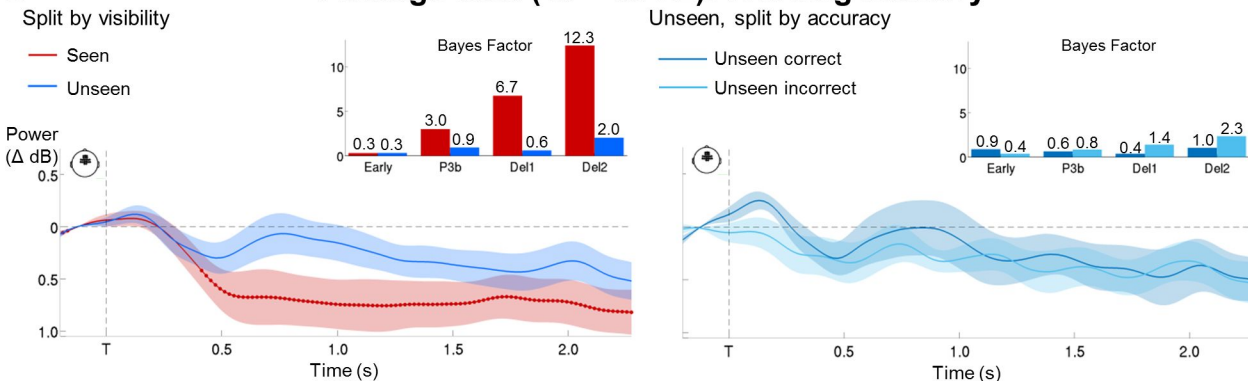
## Decoding visibility/accuracy in the alpha band (8 – 12 Hz)



**C**

## Decoding visibility/accuracy in the beta band (13 – 30 Hz)

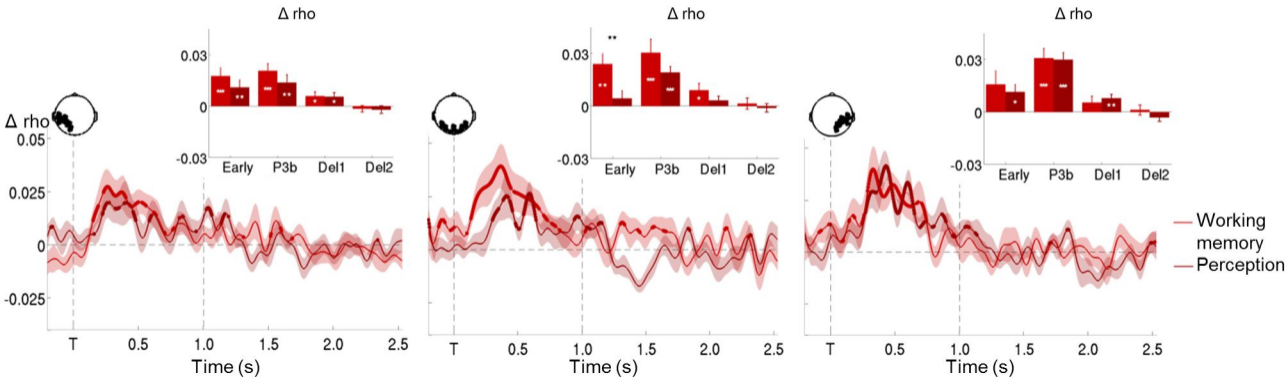


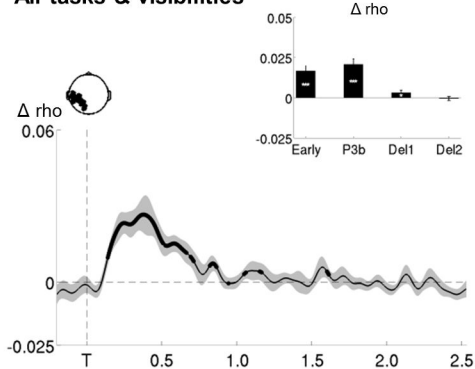
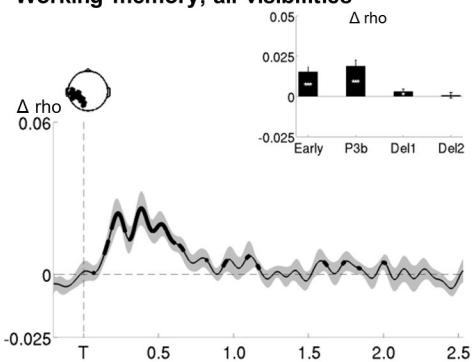
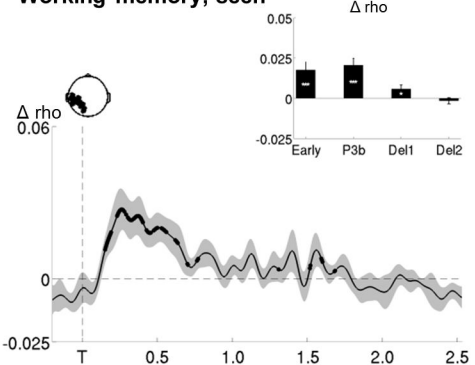
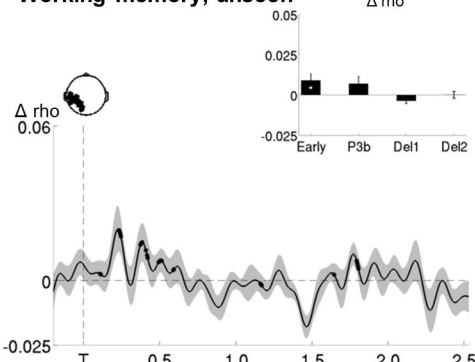
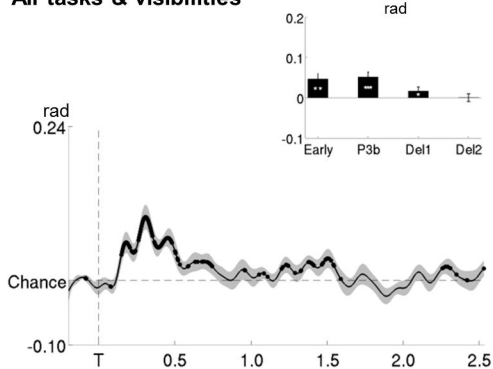
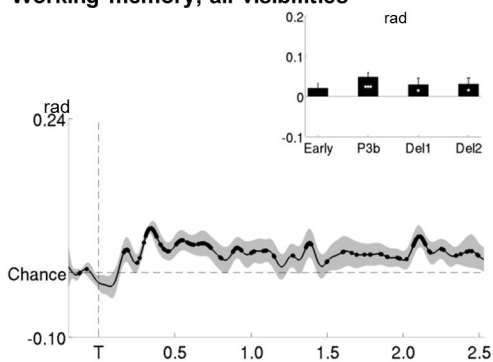
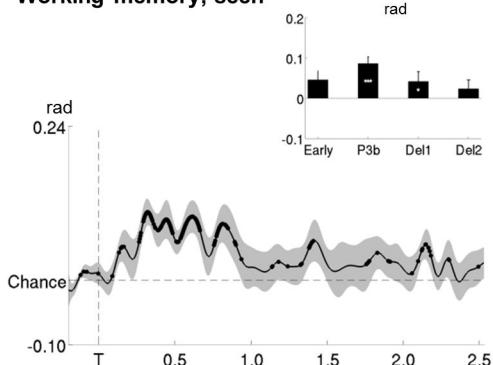
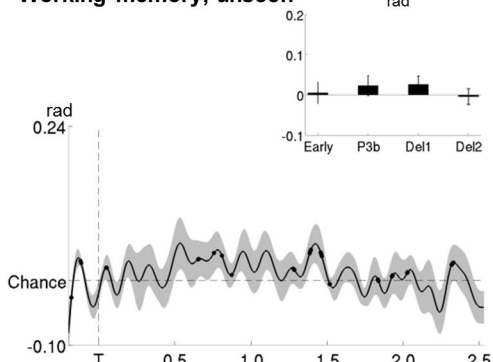
**Figure 4 – Figure Supplement 3****A****Average alpha (8 – 12 Hz): Working memory****B****Average beta (13 – 30 Hz): Working memory**



**Figure 5 – Figure Supplement 1**

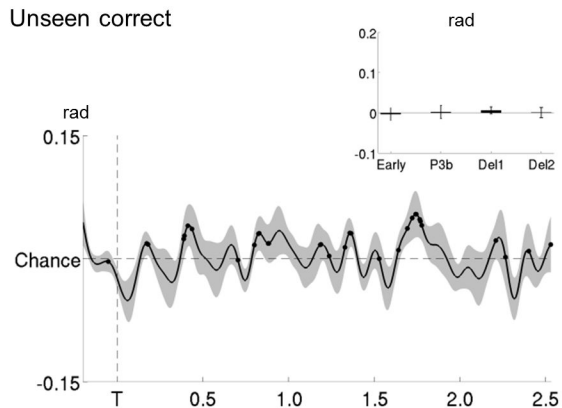
**Encoding of seen target locations in both tasks**



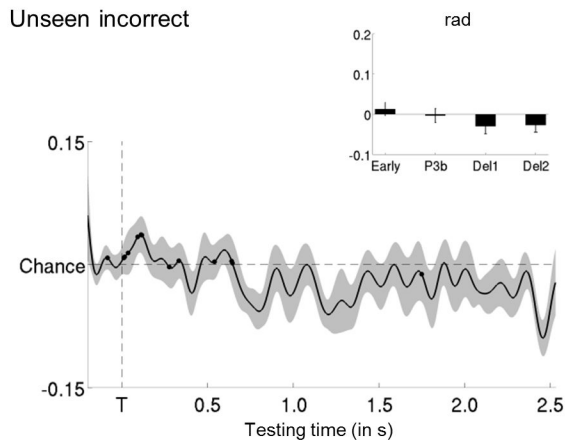
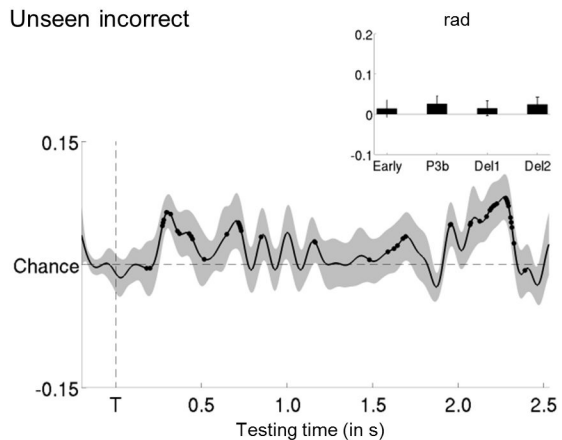
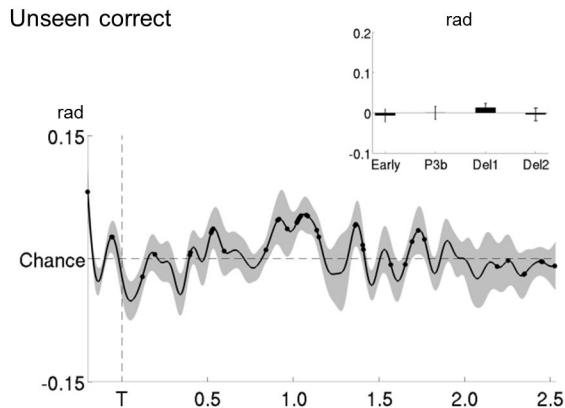
**Figure 5 – Figure Supplement 2****A Circular-linear correlations with target location****All tasks & visibilities****Working memory, all visibilities****Working memory, seen****Working memory, unseen****B Decoding target location****All tasks & visibilities****Working memory, all visibilities****Working memory, seen****Working memory, unseen**

## Figure 5 – Figure Supplement 3

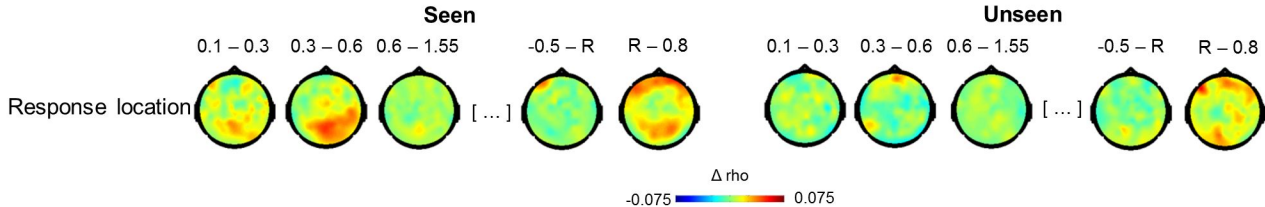
### A Decoding target location (trained on seen correct)



### B Decoding response location (trained on seen correct)



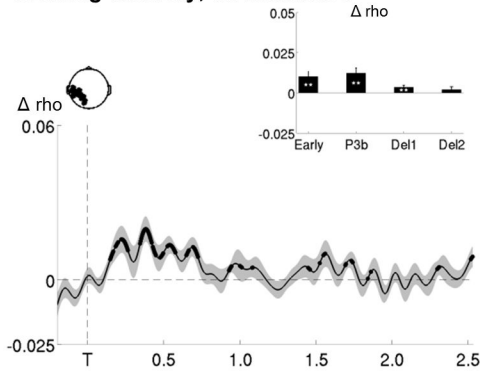
# Figure 6 – Figure Supplement 1



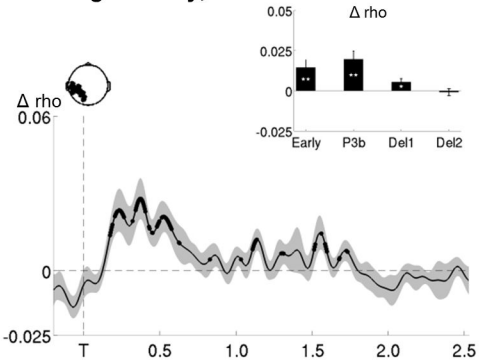
# Figure 6 – Figure Supplement 2

## A Circular-linear correlations with response location

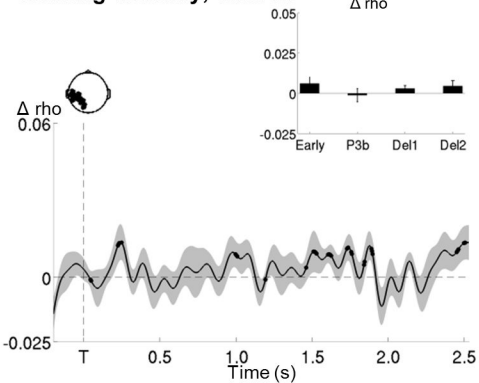
Working memory, all visibilities



Working memory, seen

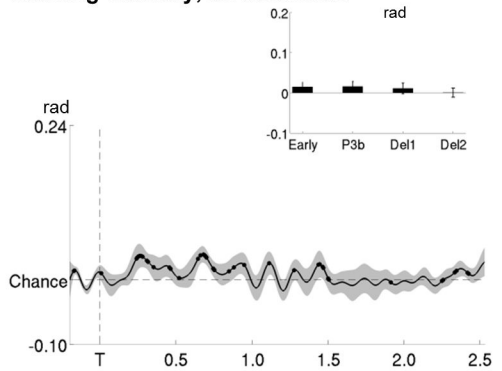


Working memory, unseen

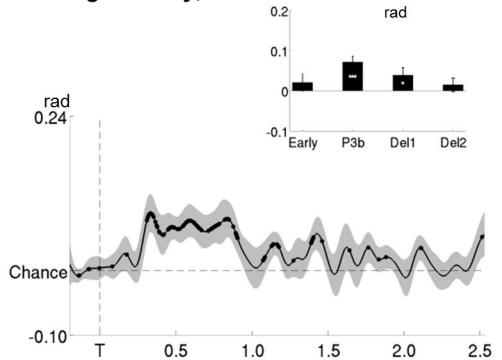


## B Decoding response location

Working memory, all visibilities



Working memory, seen



Working memory, unseen

