nature neuroscience

BRIEF COMMUNICATION

https://doi.org/10.1038/s41593-018-0094-4

Cortical specialization for attended versus unattended working memory

Thomas B. Christophel[1,6*], Polina Iamshchinina[1,6], Chang Yan[1], Carsten Allefeld[1] and John-Dylan Haynes[1,2,3,4,5]

Items held in working memory can be either attended or not, depending on their current behavioral relevance. It has been suggested that unattended contents might be solely retained in an activity-silent form. Instead, we demonstrate here that encoding unattended contents involves a division of labor. While visual cortex only maintains attended items, intraparietal areas and the frontal eye fields represent both attended and unattended items.

The short-term retention of sensory stimuli in working memory is fundamental to human cognition[1]. A wide range of primate electrophysiology and human imaging studies have reported content-selective brain signals that encode working memory contents across brief delays[2]. Such persistent, stimulus-selective activity has been observed in multiple regions across the cortical sheet, including sensory, parietal and frontal regions[2]. Recently, however, it has been postulated that working memory can be retained in an 'activity-silent state'[3–5]. In this model, working memory contents are believed to be retained by changes in synaptic weights rather than neuronal firing[3,4]. In line with this, several studies have recently reported absence of persistent stimulus-selective activity when items are held in memory, but are currently not behaviorally relevant[6–8]. Such currently nonprioritized items are frequently referred to as 'unattended memory items' (UMIs), as opposed to 'attended memory items' (AMIs)[5]. These results suggest that attended memory items are retained actively while unattended memory items are retained in an activity-silent form.

However, the absence of content-selective signals for unattended items observed in prior work[6–8] might reflect a lack of sensitivity in the experimental procedures. For example, these studies used small numbers of subjects, trained their classification models on attended items in separate one-item tasks and only analyzed limited sets of voxels or electrodes, leaving it possible that unattended items might be represented in other brain areas or using an orthogonal neural code[9]. Here we test directly whether brain regions in not only sensory but also parietal and frontal cortex contain memory representations during the delay phase for unattended stimuli. We acquired functional MRI data from a large pool of subjects ($n=87$) while they were memorizing orientation stimuli (Supplementary Fig. 1). We used a working memory design that allowed us to separately identify representations of attended and unattended stimuli. In each trial, participants first memorized the orientation of two gratings (Fig. 1). After presentation of these stimuli, a retrocue indicated which of the two gratings would be tested in an upcoming change discrimination task following an extended delay, which is the main retention interval in our design. Then, after this memory test, a second retrocue was shown that selected either the same or the other orientation for a second memory test. Such a two-stage retention task forces participants to maintain the orientations of both gratings until the second retrocue, but prioritizes and thus directs attention to the first retrocued item (AMI) while minimizing attention on the other item (UMI)[6,7].

We used a variant of multivariate pattern analysis (cvMANOVA; see Methods for details)[10] to identify which brain regions encoded the memorized orientations for attended and unattended items. The experiment was designed to optimize the ability to detect memory information in the main retention interval following the first retrocue (Fig. 1 and see Methods for details). We analyzed stimuli in each hemifield separately to account for differences in retinotopic location. Our analysis focused on the set of regions where prior work indicated the presence of persistent stimulus-selective activity for orientations when attention was not manipulated[11–13]: visual cortex (V1–V4), intraparietal sulcus (IPS0–IPS5) and the frontal eye fields (FEF; Fig. 2a).

In early visual cortex, we found reliable information about attended memory items (Fig. 2b; one-tailed one-sample t test; $t_{86}=3.37$, $P=0.000558$) whereas we found no significant information for unattended items (one-tailed one-sample t test; $t_{86}=0.19$, $P=0.423091$, lower 95th percentile confidence interval, corrected ($CI^{95}{}_{corrected}$)$=0.01$). Information was also significantly higher for attended than for unattended items (two-tailed paired-sample t test; $t_{86}=2.65$, $P=0.009467$, lower $CI^{95}{}_{corrected}=-0.012$, difference in mean pattern distinctness (ΔD) $CI^{95}{}_{corrected}=[0.007, 0.048]$). This finding closely resembles previous reports that unattended memory items are not accompanied by delay-period information in perceptually driven brain regions[7]. Our data, as shown in Supplementary Fig. 2, suggests that this attention effect is primarily driven by V1. It is worth noting that we cannot exclude the possibility that more sensitive methods might reveal information for unattended items also in visual cortex. Furthermore, whether attended or unattended items can be decoded in the current study might depend on using a larger sample size than in prior work (Supplementary Fig. 3).

Regardless, if our analyses had focused exclusively on these visual brain regions, we might have concluded that working memory representations for unattended stimuli are silent during the delay.

[1]Bernstein Center for Computational Neuroscience and Berlin Center for Advanced Neuroimaging and Clinic for Neurology, Charité Universitätsmedizin, corporate member of Freie Universität Berlin, Humboldt Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany. [2]Berlin School of Mind and Brain, Humboldt Universität, Berlin, Germany. [3]Cluster of Excellence NeuroCure, Charité Universitätsmedizin, corporate member of Freie Universität Berlin, Humboldt Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany. [4]Department of Psychology, Humboldt Universität zu Berlin, Berlin, Germany. [5]SFB 940 Volition and Cognitive Control, Technische Universität Dresden, Dresden, Germany. [6]These authors contributed equally: Thomas B. Christophel and Polina Iamshchinina. *e-mail: tbchristophel@gmail.com

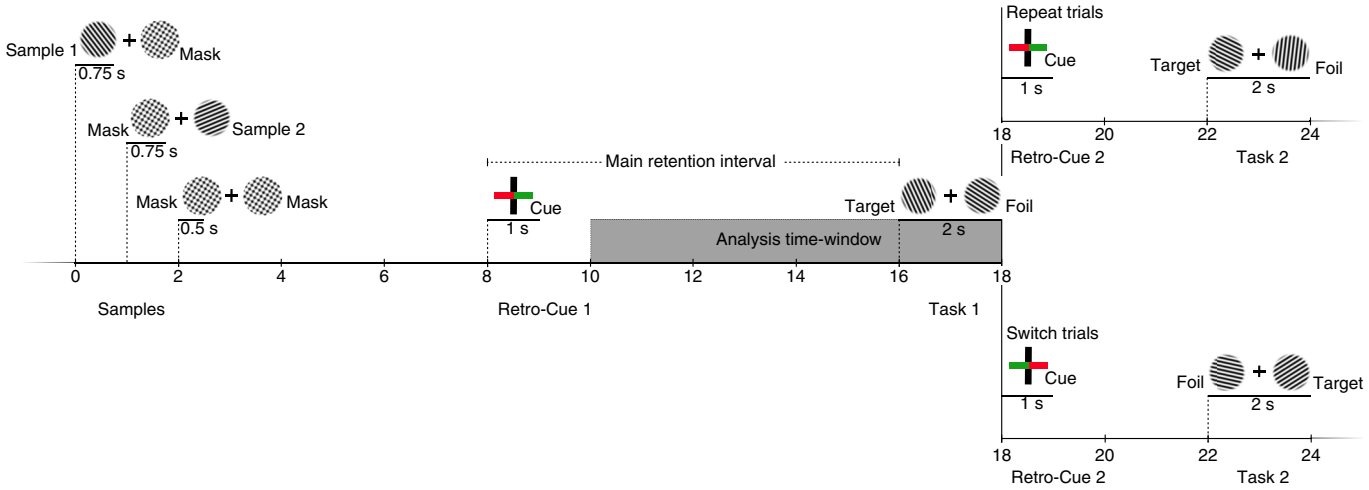

Fig. 1 | Two-stage orientation-change discrimination task using retrocue selection. In each trial, subjects are first presented with two sequential memory displays (sample 1 and sample 2; Supplementary Fig. 1). The first presents one memory item (a Gabor patch of varying orientation) on one side and a cross-hatched plaid mask on the opposite side. The second display presents a second memory item on the other side, again accompanied by a mask on the opposite side. The sequential presentation was chosen to avoid perceptual grouping of both memory items. The two memory displays were followed by a screen with cross-hatched backward masks in the previous locations of the stimuli. After a 5.5-s delay the first retrocue (red) indicated the side of the sample orientation that should be used for the first upcoming change discrimination task. This was followed by the main retention interval of 8 s. Following the main retention interval, participants viewed a Gabor patch presented on the cued (red) side and were required to judge whether it was rotated clockwise or counterclockwise compared to the cued sample. A random foil orientation was shown on the noncued (green) side. This was followed by a second retrocue, which indicated either the same (i.e., the previously attended) memory sample (repeat trials) or the other (previously unattended) memory item (switch trials). Following a short delay of 4 s, participants were again probed with a test item and had to perform the same orientation judgment. Thus, to solve the task, subjects had to memorize both items during the main retention interval (following retrocue 1), but one item was prioritized (AMI) over the other (UMI). Multivariate pattern analyses focused on the main retention period from 2 to 10 s after the retrocue to account for hemodynamic delays.

Similarly, when we used no anatomical constraints on the voxels used by the classification algorithm but focused on voxels activated by perception of the samples (primarily found in sensory cortices, as in ref. [7]), we found information for attended items only (one-tailed one-sample *t* test; AMI: $t_{86}=4.51$, $P=0.00001$, lower $CI^{95}=0.024$; UMI: $t_{86}=1.55$, $P=0.061992$, lower $CI^{95}=-0.0008$; two-tailed paired-sample *t* test; AMI versus UMI: $t_{86}=2.49$, $P=0.014547$, ΔD $CI^{95}_{corrected}=[0.006, 0.049]$), replicating prior work[7]. However, when we focused our analysis on anterior regions (that were not selectively tested in prior work) the picture changed. In the intraparietal sulcus and the frontal eye fields, we found that both AMIs (one-tailed one-sample *t* test; IPS: $t_{86}=3.66$, $P=0.000216$, lower $CI^{95}_{corrected}=0.012$; FEF: $t_{86}=2.53$, $P=0.006667$, lower $CI^{95}_{corrected}=0.002$) and UMIs (one-tailed one-sample *t* test; IPS: $t_{86}=3.24$, $P=0.000848$, lower $CI^{95}_{corrected}=0.005$; FEF: $t_{86}=3.81$, $P=0.000129$, lower $CI^{95}_{corrected}=0.005$) were significantly represented by neural activity patterns (Fig. 2b). We found no significant differences between information about attended and unattended items in these anterior regions (two-tailed paired-sample *t* test; IPS: $t_{86}=1.52$, $P=0.132059$, ΔD $CI^{95}_{corrected}=[-0.004, 0.033]$; FEF: $t_{86}=0.11$, $P=0.914863$, ΔD $CI^{95}_{corrected}=[-0.011, 0.014]$). This pattern of results was reflected by significant differences in the modulation factor for attention (D_{AMI}/D_{UMI}; V1–V4: 23.96; IPS: 1.98; FEF: 1.05) in early visual areas as compared to those for IPS and FEF (bootstrap confidence intervals; (V1–V4) – IPS: $CI^{95}_{corrected}=[3.1, 6.1\cdot\times10^5]$, (V1–V4) – FEF: $CI^{95}_{corrected}=[4.0, 4.7\cdot\times10^5]$, IPS – FEF [–1.4, 6.2]).

In early visual cortex, the time-course of information closely resembled that found by prior work[6,7], showing null results for unattended items 2 s after cue onset, whereas more anterior areas appear to represent unattended items as late as 6–10 s after the cue (Supplementary Fig. 4). To explore this further, we asked whether similar brain activity patterns represented the remembered items in data recorded in the 2 s before the cue and in the time period 6–10 s after the cue. We found such pattern stability, *xD* (see Methods for details), across time only in the IPS and FEF (one-sided one-sample *t* test; early visual cortex: $xD_{AMI}=0.0020$, $t_{86}=1.24$, $P=0.1088$; $xD_{UMI}=0.0017$, $t_{86}=0.90$, $P=0.1845$; IPS: $xD_{AMI}=0.0063$, $t_{86}=3.21$, $P=0.0009$; $xD_{UMI}=0.0027$, $t_{86}=1.51$, $P=0.0678$; FEF: $xD_{AMI}=0.0031$, $t_{86}=2.70$, $P=0.0042$; $xD_{UMI}=0.0050$, $t_{86}=3.64$, $P=0.0002$).

Finally, we tested whether voxels in the hemisphere contralateral to sample presentation carry more information about these memorized contents than ipsilateral voxels. In visual cortex, consistent with prior work[14], we found no evidence for lateralization (one-tailed paired-sample *t* test; AMI: $t_{86}=-1.13$, $P=0.8696$; UMI: $t_{86}=-1.60$, $P=0.9437$). The FEF showed similar results (AMI: $t_{86}=-0.11$, $P=0.5447$; UMI: $t_{86}=1.11$, $P=0.1341$). In intraparietal areas, however, we found more information regarding attended items in contra- versus ipsilateral voxels (AMI: $t_{86}=2.84$, $P=0.0027$; UMI: $t_{86}=-2.1$, $P=0.9823$), and contralateral areas carried more information about attended than unattended items (one-tailed paired-sample *t* test; $t_{86}=4.55$, $P=0.000009$).

Our results directly contradict the assertion that unattended working memory items are encoded solely in an activity-silent fashion. We cannot discern whether stimulus-selective persistent activity represents an active recurrent excitation network[15] or selective activity related to other potential forms of retention. Current computational models of retention via synaptic plasticity, for example, either (i) require neuronal firing as a means to uphold synaptic signals over longer periods of time[4] or (ii) suggest a complimentary role of recursive activity and synaptic plasticity[16]. Selective changes in synaptic plasticity could even lead to purely epiphenomenal selective firing. Thus, the presence or absence of stimulus-selective persistent activity[2] (in spiking, local field potentials or blood-oxygen-level dependent (BOLD) activity) does not rule out synaptic contributions to working memory.

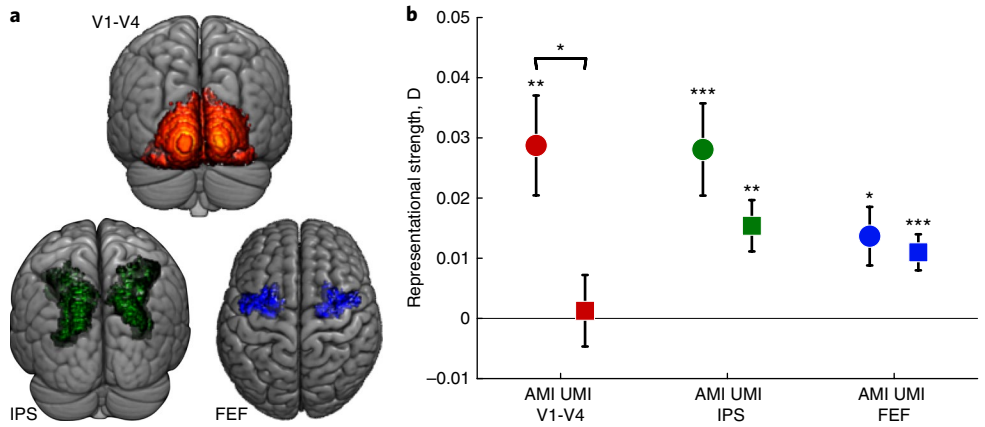

Fig. 2 | Representation of attended and unattended memory items. a, Rendered representations of the human brain depicting the three main regions of interest (ROIs): visual areas (V1–V4; red), intraparietal areas (IPS0–IPS5; green) and FEF (blue). **b**, Information about AMIs (circles) and UMIs (squares) as indicated by mean pattern distinctness, *D*, within each ROI ($n = 87$ human subjects; error bars indicate s.e.m.; one-tailed one-sample *t* tests and two-tailed paired-sample *t* tests; *P 0.05; **$P < 0.01$; ***$P < 0.001$; all Bonferroni-corrected for multiple comparisons).

Critically, our results do provide support for a different hypothesis explaining how attended and unattended items differ in their neural representation. One possibility is that sensory cortex maintains a high-resolution representation of the currently attended memory item, whereas parietal cortex has low-resolution representations of both attended and unattended items[2,9]. If that were the case, one would expect working memory performance to be more accurate for previously attended items. In line with this, behavioral evidence shows that retention in an unattended state can result in impaired change detection[17], more guesses and nontarget responses[18], stronger categorical biases[19] and less precision[18,19] as compared to attended-memory items. Behavioral data from the current study are consistent with these findings (Supplementary Fig. 5). We found selective recruitment of early visual cortex for the retention of attended memories, which could be the neural source of these behavioral benefits. In line with this, imaging evidence shows that the precision of neural representations in visual cortex during the delay period correlates with behavioral precision during recall[20]. In the current study, the amount of information we found in early visual cortex for attended memory items correlated with the individual subjects' discrimination threshold (Pearson's linear correlation coefficient; $r = -0.3177$, $P = 0.0027$).

We thus propose that the attentional modulation of working memory is not implemented by switching from an activity-based to an activity-silent synaptic code or by increasing the level of selective activity for one representation globally across all areas. Instead, attentional prioritization in working memory might be realized by selective recruitment of sensory representations that more precisely retain the information for an upcoming task.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41593-018-0094-4.

Received: 7 July 2017; Accepted: 23 January 2018;
Published online: 05 March 2018

References

1. Baddeley, A. D. *Working Memory* (Clarendon Press, Oxford, 1986).
2. Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R. & Haynes, J.-D. *Trends Cogn. Sci.* **21**, 111–124 (2017).
3. Stokes, M. G. *Trends Cogn. Sci.* **19**, 394–405 (2015).
4. Mongillo, G., Barak, O. & Tsodyks, M. *Science* **319**, 1543–1546 (2008).
5. Larocque, J. J., Lewis-Peacock, J. A. & Postle, B. R. *Front. Hum. Neurosci.* **8**, 5 (2014).
6. Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K. & Postle, B. R. *J. Cogn. Neurosci.* **24**, 61–79 (2012).
7. LaRocque, J. J., Riggall, A. C., Emrich, S. M. & Postle, B. R. *Cereb. Cortex* **27**, 4881–4890 (2017).
8. Wolff, M. J., Jochim, J., Akyürek, E. G. & Stokes, M. G. *Nat. Neurosci.* **20**, 864–871 (2017).
9. Olivers, C. N. L., Peters, J., Houtkamp, R. & Roelfsema, P. R. *Trends Cogn. Sci.* **15**, 327–334 (2011).
10. Allefeld, C. & Haynes, J.-D. *Neuroimage* **89**, 345–357 (2014).
11. Harrison, S. A. & Tong, F. *Nature* **458**, 632–635 (2009).
12. Serences, J. T., Ester, E. F., Vogel, E. K. & Awh, E. *Psychol. Sci.* **20**, 207–214 (2009).
13. Ester, E. F., Sprague, T. C. & Serences, J. T. *Neuron* **87**, 893–905 (2015).
14. Ester, E. F., Serences, J. T. & Awh, E. *J. Neurosci.* **29**, 15258–15265 (2009).
15. Hebb, D. O. *The Organization of Behavior*(Wiley, New York, 1949).
16. Itskov, V., Hansel, D. & Tsodyks, M. *Front. Comput. Neurosci.* https://doi.org/10.3389/fncom.2011.00040 (2011).
17. Rerko, L. & Oberauer, K. *J. Exp. Psychol. Learn. Mem. Cogn.* **39**, 1075–1096 (2013).
18. Emrich, S. M., Lockhart, H. A. & Al-Aidroos, N. *J. Exp. Psychol. Hum. Percept. Perform.* **43**, 1454–1465 (2017).
19. Bae, G.-Y. & Luck, S. *J. Vis.* https://doi.org/10.1167/16.12.701 (2016).
20. Ester, E. F., Anderson, D. E., Serences, J. T. & Awh, E. *J. Cogn. Neurosci.* **25**, 754–761 (2013).

Acknowledgements

This work was funded by the Bernstein Computational Neuroscience Program of the German Federal Ministry of Education and Research BMBF Grant 01GQ0411, by the Excellence Initiative of the German Federal Ministry of Education and Research and DFG Grants GSC86/1-2009, KFO247, HA 5336/1-1, SFB 940 and JA 945/3-1/SL185/1-1.

Author contributions

T.B.C., P.I. and J.-D.H. designed the study. P.I. and C.Y. acquired data. T.B.C., P.I. and C.A. analyzed the data. T.B.C., P.I., C.A. and J.-D.H. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41593-018-0094-4.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to T.B.C.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Participants. We recruited 89 healthy right-handed human subjects (42 female; mean age: 26.8, s.e.m. ± 0.4) with normal or corrected-to-normal vision for the current study. We based our estimate of the necessary sample size on our prior work on attended working memories[21–24], as effect sizes for unattended memories are unknown. We substantially increased *N* value relative to these studies to account for putative reductions in effect size for unattended items and because of a generally lower trial count. Data acquisition for two subjects was aborted during the experiment upon the request of the participants. This was before any stage of data analysis. We thus analyzed data from 87 participants (41 female; mean age: 26.8, s.e.m. ± 0.4). Subjects gave informed consent and the study was approved by the local ethics committee (Ethics committee, Department of Psychology, Humboldt University, Berlin). Data collection and analysis were not performed blind to the conditions of the experiment.

Procedure and design. In an MRI scanner, subjects performed a two-stage delayed change-discrimination task[6,7] using Gabor grating stimuli. During this task, subjects memorized two gratings and were twice instructed, using retrocues, to attend to one or the other for an upcoming change-detection task. Critically, this experimental scheme results in a situation after the first cue in which one orientation sample is prioritized (the attended memory item, AMI) while the other sample is remembered but with lower priority (the unattended memory item, UMI).

In each trial, participants viewed two sine-wave sample grating stimuli shown consecutively left and right to the center of the screen in random order (0.8° off center, for 750 ms, ISI 250 ms). During grating presentation, the side opposite the grating was occupied by a Gabor cross-hatched stimulus and the same cross-hatched stimuli were shown as masks after stimulus presentation (for 500 ms). The onset of these masks was followed, after 5,500 ms, by retrocue 1 (presented for 1,000 ms). For this cue, one of the arms of the fixation cross turned red while the opposite arm turned green. Participants were instructed that the sample grating that had been shown on the red side of the fixation cross was to be used for the upcoming change detection task.

This first retrocue was followed by a prolonged delay (7,000 ms), the main retention interval, during which participants were expected to maintain orientation representations of both sample gratings with the cued orientation prioritized over the other. This delay is the critical time-period of interest, during which neural representations of prioritized contents (attended memory items, AMI) and nonprioritized information (unattended memory items, UMI) can be distinguished. During the change discrimination task that followed (task 1, 2,000 ms), participants had to report whether a target grating presented on the cued side was rotated clockwise or counterclockwise relative to the cued sample grating previously presented on the same side. On the opposite side of the screen a randomly rotated foil grating was shown.

This first task was followed by a second cue (retrocue 2, 1,000 ms), an additional delay (3,000 ms) and a second task (task 2, 2,000 ms). The grating cued for the second task could be either the same as for the first task (repeat trials) or the other grating (switch trials). The switch probability was 50%. This second task ensured that an item not prioritized for the first task needed to be retained until the second cue was presented because subjects did not know in advance whether the item might be relevant later. Please note that our study was designed to maximize sensitivity in the first delay period. For this reason, the second delay was chosen to be very short. Thus, we did not perform an analysis of a potential reinstatement of information in the 50% switch trials when a stimulus feature switched from unattended to attended[6,7]. This was further motivated by recent proposals that evidence of reinstatement is not conclusive evidence of silent working memory[25]. The intertrial interval was either 2,000 ms (50% of trials), 4,000 ms (33.3%) or 6,000 ms (16.7%). A fixation cross (width 0.2°) remained on screen throughout the experiment.

Prior to the main experiment in the MRI scanner (2–4 d in advance), participants took part in a training session outside the scanner using a conventional LCD display. The training comprised four experimental runs with shorter delays (1,500 ms and 2,500 ms for the first and the second delay periods, respectively) and using fully randomized sample orientations. Subjects were instructed not to use verbal labels for memorizing the stimuli. In the MRI scanner, the experimental scheme was presented on a NordicNeuroLab Monitor (70.5 cm wide) and subjects viewed the screen via a mirror. Stimulus presentation was controlled using Psychtoolbox[26].

The stimuli used were sine-wave Gabor gratings (5.7° size, phase-randomized, spatial frequency: 1.8 cycles/degree, 3.06° from the screen center) with varying orientations. Notably, sample orientations presented (Supplementary Fig. 1) in the left visual field were drawn from a different pool (7.5°, 37.5°, 67.5°, 97.5°, 127.5°, 157.5°) than orientations on the right (22.5°, 52.5°, 82.5°, 112.5°, 142.5°, 172.5°), which allowed us to decorrelate items shown in the left and right hemifield and thereby AMIs and UMIs[27]. Gabor cross-hatched plaid stimuli (5.7° size) consisting of two random but orthogonal orientations (phase randomized, spatial frequency: 1.8 cycles/degree) were used as masks.

For the change-discrimination tasks, targets were rotated clockwise and counterclockwise on an equal number of trials. The extent of rotation of the test grating was initially set to 20° and adjusted using a staircase procedure to generate a consistently challenging task and avoid ceiling effects. For each correct response in a given trial (0–2), the difference between test and sample orientation was reduced by 0.5°, making change discrimination harder. Conversely, the difference was increased by 2° for each incorrect response, thus making them easier to differentiate. Changes to this discrimination threshold were only applied after the end of a given trial, and the same levels were used for task 1 and task 2, to allow for comparisons between the tasks. Adjustment started during training and continued throughout the fMRI experiment.

There were four scanning runs of 48 trials each. We used a within-subject 2 (retrocue 1: left vs. right) × 2 (retrocue 2: switch vs. repeat) design. Each of the 12 orientations (six for each side) had to be memorized in 8 trials per run. The pairing of orientations on the left and right was fully randomized to allow statistically independent analyses of attended and unattended orientations. The assignment of attention conditions to orientation conditions was fully randomized in the first 39 participants. Based on theoretical considerations[28] we then decided to slightly modify the randomization so that in the second set of 48 participants each attention condition was associated with exactly the same number of trials for each orientation. We found no differences between the groups in our main analyses. For the change discrimination tasks, the rotation of the test orientation was randomized with respect to all other conditions with equal frequency of clockwise and counterclockwise rotations for each of the two tasks. The temporal order of conditions was fully randomized within each run.

Overall, the experiment took 90 min per participant. In order to decrease effects of long-term memory, a short 5-min run was presented between the second and third experimental block, in which participants performed the task with random sample orientations while anatomical scans were acquired. After the experiment, participants filled out a questionnaire covering the strategies they had used for memorization.

Data acquisition. fMRI data were collected with a Siemens 3-Tesla TIM-Trio MR tomograph located at the Berlin Center for Advanced Imaging (Charité-Universitätsmedizin). Within each of the four runs, we recorded 663 T2*-weighted gradient-echo echoplanar images (EPI, 33 slices, 3 × 3 × 3-mm resolution, 0.6-mm gap, descending order, FoV = 192 mm, TR = 2,000 ms, TE = 30 ms, flip angle = 80°). The onset of each trial was locked to the onset of image acquisition to minimize variation due to slice-acquisition onsets. Slices were aligned parallel to the anterior and posterior commissures and covered the whole neocortex. In addition, a high-resolution T1-weighted image was acquired (192 sagittal slices, 1 mm thickness, RT = 1,900 ms, TE = 2.52 ms, flip angle = 9°, FOV = 256 mm).

fMRI preprocessing. Functional imaging data was analyzed using SPM12[29] and cvMANOVA[10]. After conversion to NIfTI format, the functional data were motion corrected and the anatomical image was coregistered to the first image of the BOLD time-series. No normalization into standard space was performed and we applied no Gaussian smoothing to the data before performing the multivariate analyses, to preserve the fine-scaled spatial structure of the fMRI data. Using a similar reasoning, we abstained from using slice-time correction during preprocessing and avoided any other temporal filtering to retain the temporal precision of the stimulus-locked time-series (see the "Data acquisition" section).

Anatomical regions of interest (ROIs). We aimed to identify information about attended and unattended memory items within brain regions previously found to carry information about memorized gratings[11–13]. For this, anatomical probability maps of retinotopic areas[30] in visual cortex (V1–V4), the intraparietal sulcus (IPS0–IPS5) and the frontal eye fields (Fig. 2a) were backward-transformed into participants' native space using unified segmentation[31]. These maps were thresholded to exclude voxels with a probability of being part of a given area < 0.1. In a post hoc exploratory analysis, we investigated the information content of V1 to V4, separately. Please note that an analysis based on individual retinotopic maps might have had higher sensitivity to detect weak effects in areas beyond V1. For the main analyses, these anatomical masks were collapsed across the left and the right hemispheres. Separate masks for the left and right hemispheric portions of these areas were used to investigate the lateralization of representations in a post hoc exploratory analysis. Finally, we also performed a post hoc exploratory analysis without any anatomical preselection of voxels to compare our results to previous studies[7].

Univariate analyses of sample-related activity. To estimate BOLD activity during the trial, a GLM with seven regressors was designed using hemodynamic response functions (HRF) time-locked to the onsets of the following events and adjusted by their duration: sample grating onsets (first regressor), first and second cue onsets (second–third), each of the delay period onsets separately modeled taking into account their respective durations (fourth, before the first cue; fifth, following the first cue; sixth, following the second cue) and both

discrimination task onsets merged into one regressor (seventh). To identify voxels that responded to our grating stimuli, we generated *t*-maps contrasting the sample grating onsets (irrespective of orientation) against the implicit baseline of the model. Then, individual subject-level *t*-maps were overlaid with each of the ROIs created for each participant. To avoid an arbitrary selection of *n* voxels for each ROI, we generated 25 versions of every ROI, each representing the *n* voxels with the highest sample-related activity. For this, we varied n between 20 and 500 voxels in steps of 20 voxels. We only considered voxels that were positively activated ($t > 0$). These ROIs with varying sizes were later used in a nested cross-validation to choose an optimal voxel number for each ROI, while at the same time avoiding overfitting.

Analyses of multivariate pattern distinctness using cvMANOVA. The goal of the multivariate pattern analyses was to test whether retinotopic areas in visual, parietal and frontal cortex have representations of the remembered attended and unattended orientations. For this, we used a recently developed technique for multivoxel pattern analysis, cross-validated MANOVA[10]. cvMANOVA constitutes a variant of multivariate analyses of variance[32] that can be used to quantify differences in BOLD response patterns[24,33]. The method is comparable to more common classifier-based 'decoding' analyses[11–14,20–24,34–43] but has a number of advantages: it avoids binary classification in favor of a continuous measure of patterned differences, performs a parameter-free analysis based on a probabilistic model of the data (the multivariate general linear model) and results in an interpretable multivariate effect size (explained variance). Moreover, since *D* is a cross-validated version of a likelihood-ratio statistic, it can be expected to be more sensitive than classification accuracy (cf. Fig. 3d in ref. [10]). Please note that prior work employed data from one-item tasks (i.e., including only attended items) for classifier training to avoid training on ambiguous two-item data where the representations of two items might overlap[6,7]. This analysis potentially biases the results in favor of attended memory items. Prior work failed to identify information for either attended or unattended items when only using data from two-item tasks[7], possibly due to a lack of power. Here, to avoid biasing our results, we only recorded data in two-item tasks and substantially increased the *N* to counteract the lowered power when training on such data.

Here we used cvMANOVA to ask whether activity patterns in our ROIs carried information about the memorized contents. As a first step, a multivariate general linear model (MGLM) using finite impulse–response (FIR) functions was used to estimate memory-related activity in each voxel. Two first-level models were estimated for samples presented in the left and the right visual fields, respectively. For each of the 6 orientations per side, we used 12 regressors to model the entire 24 s of each trial in 2-s time bins (i.e., the length of the TR) spanning the time between onset of the masks (after sample presentation) until 2 s after the end of the trial (only a subset of this time was in the actual decoding analysis time window; see below and Fig. 1). This set of regressors was modeled for each orientation in two different conditions: when a particular orientation was selected by the cue and when it was not (24 regressors per sample grating). BOLD activity in each voxel was fitted with this set of FIR regressors separately for each of the four runs. Thus, we had 144 regressors per run for the left and the right side separately (factors: timepoint (12) × attended/unattended (2) × orientation (6)) collapsing across repeat and switch trials, as well as a constant regressor to model the run mean. Parameter estimates and residuals from these two models were then used in the region-based cvMANOVA.

Within each model, to test the effect of orientation identity, we used two contrast matrices, which (with respect to a single time point) had the forms

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \text{ for attended items and } \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & -1 \end{pmatrix}$$

for unattended items, where the 12 rows correspond to the 12 regressors for attended and unattended orientations.

The columns correspond to the five partial contrasts (comparing orientations 1 vs. 2, 2 vs. 3, 3 vs. 4, 4 vs. 5, and 5 vs. 6) together defining the effect of orientation. We analyzed the four timepoints following the presentation of the first cue (i.e., FIR bins 5–8, when attended and unattended items were differentiated) with a shift of 2 s to account for hemodynamic delay. For this, the two contrast matrices were replicated and zero-padded. This can be exemplified for attended memory items (A = 1) as follows:

T	O	A						
1	1	1	0	0	0	0	0	⋯
2	1	1	0	0	0	0	0	⋯
3	1	1	0	0	0	0	0	⋯
4	1	1	0	0	0	0	0	⋯
5	1	1	1	0	0	0	0	⋯
6	1	1	0	1	0	0	0	⋯
7	1	1	0	0	1	0	0	⋯
8	1	1	0	0	0	1	0	⋯
9	1	1	0	0	0	0	0	⋯
10	1	1	0	0	0	0	0	⋯
11	1	1	0	0	0	0	0	⋯
12	1	1	0	0	0	0	0	⋯
1	2	1	0	0	0	0	0	⋯
2	2	1	0	0	0	0	0	⋯
3	2	1	0	0	0	0	0	⋯
4	2	1	0	0	0	0	0	⋯
5	2	1	−1	0	0	0	1	⋯
6	2	1	0	−1	0	0	0	⋯
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋱

Here, the rows correspond to the 144 regressors per run, with FIR regressors for all 12 timepoints (T) of each orientation condition (O) grouped together. For the two contrasts per model we estimated the amount of multivariate variance explained by the encoded effect relative to the multivariate error variance using cross-validated MANOVA[10].

cvMANOVA estimates the variance of the multivariate fMRI time-series that can be explained by a contrast between conditions (here: differences in multivariate responses to different orientations). The pattern distinctness is

$$D = trace\left(\frac{1}{n} B' C C^{-} X' X C C^{-} B \cdot \Sigma^{-1}\right)$$

where *C* is the contrast matrix, *B* is the parameter matrix, n is the number of scans, *X* is the design matrix, and Σ is the error covariance matrix. *D* is the amount of multivariate variance explained by the effect encoded in the contrast, in units of the multivariate error variance. To obtain an unbiased estimate of the explained variance cross-validation is used (see ref. [10] for details). Because the number of voxels in a region may be large in some cases, explained variance was computed relative to an estimate of the multivariate error variance, Σ, which was regularized toward the diagonal, using an optimized regularization parameter[44].

In the present application, if different orientations elicit the same multivariate response, *D* would on average be 0, while different responses to different orientations would lead to an average *D* larger than 0. *D* for attended and unattended orientations was averaged across orientations shown in the left and right visual field.

This procedure was performed separately for all ROIs using varying voxel counts (between 20 and 500 voxels, see above). To select the optimal number of voxels for each area within each subject and analysis while avoiding double-dipping, we used a nested cross-validation approach. For every subject, we averaged *D* across all other subjects for all 25 possible ROI sizes and selected the ROI size with maximal *D*. *D* for the left-out subject with this ROI size was kept and this procedure was repeated for every subject.

While the main analysis used one contrast for attended and unattended contents, each across timepoints, we also performed a post hoc exploratory analysis that used separate contrasts for the twelve time points we estimated in the (FIR-based) multivariate general linear model to generate time courses (Supplementary Fig. 4). Above-zero *D* for each timepoint indicates that the multivariate responses to different orientations are different at that timepoint, but not necessarily that this multivariate response difference stays the same over time. We thus finally wanted to assess whether the informative brain patterns were stable across time (akin to generalization in a cross-classification analysis), in particular between the timepoint directly before the onset of the cue and the two timepoints at 6–10 s after the cue onset. Where standard cvMANOVA quantifies the pattern information in the form of variance explained by a contrast (here between different orientations), for this purpose we used a variant that estimates the amount of similarly encoded pattern information in the form of explained variance shared between two contrasts (orientation-specific responses at two different timepoints), the pattern stability

$$xD = trace\left(\frac{1}{n} B' C_1 C_1^{-} X' X C_1 C_2^{-} B \cdot \Sigma^{-1}\right)$$

where C_1 and C_2 are the two different contrast matrices, *B* is the parameter matrix, *n* is the number of scans, *X* is the design matrix, and Σ is the error covariance matrix. Again, cross-validation is used to obtain an unbiased estimate. *xD* is on average 0 if there is no shared variance between the two contrasts, i.e., if the respective informative patterns are orthogonal to each other.

Statistical testing. Group-level statistics ($n = 87$ human subjects) were performed using one-sample and paired t tests. Please note that one-sample t tests do not provide population inference[45]. The data was tested for deviation from normality using Kolmogorov–Smirnov tests. We applied Bonferroni correction to all resulting P-values to account for the number of areas tested in a given analysis.

To test whether the reduction of representational strength in UMI vs. AMI is different between regions, one would normally test for an interaction using an ANOVA with factors attention and region. Standard interaction analyses test the difference (for example, between areas) of a difference (for example, between the conditions, $D_{AMI} - D_{UMI}$) for significance. However, information measures are generally not comparable between brain areas as a result of their unique neural topology, vascular structure and levels of physiological noise[46–48]. In a post hoc exploratory analysis, we therefore adopted a strategy of first quantifying the effect of attention within region by an attentional modulation factor D_{AMI}/D_{UMI}, i.e., the ratio between the average D values under AMI and UMI, which should be comparable between regions. We then computed the three pairwise differences of attenuation factors between the three regions and assessed significant difference from 0 by means of bootstrap confidence intervals on these differences[49] based on 100,000 resamples of the 87 subjects, at 95% confidence corrected for multiple comparisons.

For behavioral analyses, proportions of correct responses were calculated, treating missed responses as errors. Reaction-time analyses were conducted excluding missed responses. Differences in proportions of correct responses were tested using two-tailed Wilcoxon signed-rank tests ($n = 87$). Discrimination thresholds were averaged across all trials. Correlations between discrimination thresholds and measures of pattern distinctness for attended and unattended memory items were tested using Pearson's linear correlation coefficients.

Life Sciences Reporting Summary. Further information on experimental design is available in the Life Sciences Reporting Summary.

Code availability. Matlab source code for cvMANOVA is available online at https://github.com/allefeld/cvmanova/releases. For the analyses in this paper we used v2 (2015–1–12).

Data availability. The MRI and behavioral data that were used in this study are available to researchers from the corresponding author upon request.

References

21. Christophel, T. B., Hebart, M. N. & Haynes, J.-D. *J. Neurosci.* **32**, 12983–12989 (2012).
22. Christophel, T. B. & Haynes, J.-D. *Neuroimage* **91**, 43–51 (2014).
23. Christophel, T. B., Cichy, R. M., Hebart, M. N. & Haynes, J.-D. *Neuroimage* **106**, 198–206 (2015).
24. Christophel, T. B., Allefeld, C., Endisch, C. & Haynes, J.-D. *Cereb. Cortex* https://doi.org/10.1093/cercor/bhx119 (2017)
25. Schneegans, S. & Bays, P. M. *J. Cogn. Neurosci.* **29**, 1977–1994 (2017).
26. Brainard, D. H. *Spat. Vis.* **10**, 433–436 (1997).
27. Pratte, M. S. & Tong, F. *J. Vis.* **14**, 22 (2014).
28. Görgen, K., Hebart, M.N., Allefeld, C. & Haynes, J.-D. *Neuroimage* https://doi.org/10.1016/j.neuroimage.2017.12.083 (2017).
29. Friston, K. J. et al. *Hum. Brain Mapp.* **2**, 189–210 (1994).
30. Wang, L., Mruczek, R. E. B., Arcaro, M. J. & Kastner, S. *Cereb. Cortex* **25**, 3911–3931 (2015).
31. Ashburner, J. & Friston, K. J. *Neuroimage* **26**, 839–851 (2005).
32. Timm, N. H. *Applied Multivariate Analysis* (Springer, New York, 2002).
33. Guggenmos, M., Wilbertz, G., Hebart, M. N. & Sterzer, P. *eLife* **5**, e13388 (2016).
34. Haynes, J. D. & Rees, G. *Nat. Neurosci.* **8**, 686–691 (2005).
35. Haynes, J.-D. & Rees, G. *Curr. Biol.* **15**, 1301–1307 (2005).
36. Haynes, J.-D. & Rees, G. *Nat. Rev. Neurosci.* **7**, 523–534 (2006).
37. Kamitani, Y. & Tong, F. *Nat. Neurosci.* **8**, 679–685 (2005).
38. Soon, C. S., Brass, M., Heinze, H. J. & Haynes, J. D. *Nat. Neurosci.* **11**, 543–545 (2008).
39. Sterzer, P., Haynes, J.-D. & Rees, G. *J. Vis.* **8**, 10 (2008).
40. Cichy, R. M., Chen, Y. & Haynes, J. D. *Neuroimage* **54**, 2297–2307 (2011).
41. Riggall, A. C. & Postle, B. R. *J. Neurosci.* **32**, 12990–12998 (2012).
42. Lee, S.-H., Kravitz, D. J. & Baker, C. I. *Nat. Neurosci.* **16**, 997–999 (2013).
43. Bettencourt, K. C. & Xu, Y. *Nat. Neurosci.* **19**, 150–157 (2016).
44. Schäfer, J. & Strimmer, K. *Stat. Appl. Genet. Mol. Biol.* https://doi.org/10.2202/1544-6115.1175 (2005).
45. Allefeld, C., Görgen, K. & Haynes, J.-D. *Neuroimage* **141**, 378–392 (2016).
46. Dubois, J., de Berker, A. O. & Tsao, D. Y. *J. Neurosci.* **35**, 2791–2802 (2015).
47. Haynes, J.-D. *Neuron* **87**, 257–270 (2015).
48. Hebart, M. N. & Baker, C. I. *Neuroimage* https://doi.org/10.1016/j.neuroimage.2017.08.005 (2017).
49. Efron, B. & Tibshirani, R. J. *An Introduction to the Bootstrap* (CRC Press, Boca Raton, FL, USA, 1994).

nature research

Corresponding author(s): Christophel

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.

"We based our estimate of the necessary sample size on our prior work on attended working memories [33–36] as effect sizes for unattended memories are unknown. We substantially increased the N relative to these studies, to account for putative reductions in effect size for unattended items and because of a generally lowered trial count."

We directly adress the question of statistical power in Supplementary Figure S3.

2. Data exclusions

Describe any data exclusions.

Two subjects were excluded because data acquisition was incomplete. Incomplete data is incompatible with the cross-validation procedures used in the current study.

3. Replication

Describe whether the experimental findings were reliably reproduced.

No replication was attempted.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

"The pairing of orientations on the left and right was fully randomized to allow statistically independent analyses of attended and unattended orientations."

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Data collection and analysis were not performed blind to the conditions of the experiment. Blinding is not a common procedure for this type of research.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	Imaging data was analyzed using SPM12 and cvMANOVA.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	No unique materials were used.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	No antibodies were used.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	No eukaryotic cell lines were used.
b. Describe the method of cell line authentication used.	Not applicable
c. Report whether the cell lines were tested for mycoplasma contamination.	Not applicable
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	Not applicable

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	No animals were used in the study.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	89 healthy right-handed human subjects (42 female; mean age: 26.8, SEM ±0.4) with normal or corrected-to-normal vision were recruited for the current study.

nature research

Corresponding author(s): Christophel

☐ Initial submission ☐ Revised version ☒ Final submission

MRI Studies Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

Experimental design

1. Describe the experimental design.	slow event-related
2. Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.	87 subjects, 4 runs per subjects, 48 trials per run. 16 704 trials overall. Duration of one trial = 24 s; average ITI = 3.33 s (jitter range = 2-6 s)
3. Describe how behavioral performance was measured.	There were two tasks in each trial. For each trial, we recorded the proportion of correct responses and the response time, showing retro-cuing effects consistent with prior work (Wilcoxon signed rank tests, see Figure S3). The difficulty was adjusted after each trial in an ongoing fashion using a staircase procedure. For each trial, we recorded the state of this staircase procedure which is referred to as the discrimination threshold.

Acquisition

4. Imaging	
a. Specify the type(s) of imaging.	functional and structural MRI
b. Specify the field strength (in Tesla).	3 Tesla
c. Provide the essential sequence imaging parameters.	Within each of the four runs, we recorded 663 T2*-weighted gradient-echo echo-planar images (EPI, 33 slices, 3 × 3 × 3 mm resolution, 0.6 mm gap, descending order, FoV = 192 mm, TR = 2000 ms, TE = 30 ms, flip angle = 80°). In addition, a high-resolution T1-weighted image was acquired (192 sagittal slices, 1 mm thickness, RT = 1900 ms, TE = 2.52 ms, flip angle = 9°, FOV = 256 mm).
d. For diffusion MRI, provide full details of imaging parameters.	not applicable
5. State area of acquisition.	Slices were aligned parallel to the anterior and posterior commissures and covered the whole neocortex.

Preprocessing

6. Describe the software used for preprocessing.	Imaging data was analyzed using SPM12 and cvMANOVA.
7. Normalization	
a. If data were normalized/standardized, describe the approach(es).	We estimated normalization parameters using unified segmentation and used these parameters to warp anatomical templates from MNI into single subject space.
b. Describe the template used for normalization/ transformation.	ICBM152
8. Describe your procedure for artifact and structured noise removal.	motion correction, run-wise constant regressors, jittering (see 2).

9. Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.	No volume censoring performed.

Statistical modeling & inference

10. Define your model type and settings.	Single-subject level: Multivariate GLM, FIR regressors, no temporal filtering Between-subject level: Random effects analysis (t-test), but see discussion in Allefeld et al. (2016)
11. Specify the precise effect tested.	We tested whether - during the delay following the initial retro-cue - the six different stimulus conditions (six orientations per visual hemifield) evoked differential neural activity patterns. This hypothesis was tested seperately for cued and not cued items using a crossvalidated multivariate analysis of variance (cvMANOVA).
12. Analysis	
a. Specify whether analysis is whole brain or ROI-based.	ROI-based
b. If ROI-based, describe how anatomical locations were determined.	Anatomical locations are based on probabilistic maps of retinotopic regions of the human brain (Wang et al., 2015)
13. State the statistic type for inference. (See Eklund et al. 2016.)	ROI-wise inference
14. Describe the type of correction and how it is obtained for multiple comparisons.	Bonferroni correction
15. Connectivity	
a. For functional and/or effective connectivity, report the measures of dependence used and the model details.	not applicable
b. For graph analysis, report the dependent variable and functional connectivity measure.	not applicable
16. For multivariate modeling and predictive analysis, specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.	Independent variable: Currently remembered orientation (cued or not cued). Feature selection: Voxels within each roi were ordered by T-values indicating the contrast 'sample related activity > baseline'. The number of voxels for each analysis was selected using nested cross-validation. Model: See above, 'Single-subject level' Evaluation: We used 4-fold cross-validation leaving data from one run out for each cross-validation fold. The metric was 'pattern distinctness D' (see Allefeld & Haynes, 2014).