Working memory and imagery in early visual cortex

Simon Weber[a,b], Thomas Christophel[a,c], Kai Görgen[a,d],

Joram Soch[a,e], John-Dylan Haynes[a,b,c,d,f,g].

a Bernstein Center for Computational Neuroscience Berlin and Berlin Center for Advanced Neuroimaging, Charité Universitätsmedizin Berlin, corporate member of the Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany
b Research Training Group "Extrospection" and Berlin School of Mind and Brain, Humboldt-Universität zu Berlin, Berlin, Germany
c Department of Psychology, Humboldt-Universität zu Berlin, Berlin, Germany
d Research Cluster of Excellence "Science of Intelligence", Technische Universität Berlin, Berlin, Germany
e German Center for Neurodegenerative Diseases (DZNE), Göttingen, Germany
f Collaborative Research Center "Volition and Cognitive Control", Technische Universität Dresden, Dresden, Germany
g Max Planck School of Cognition, Stephanstrasse 1a, Leipzig, Germany

Keywords

Working memory, visual imagery, early visual cortex, individual differences, multivariate decoding

Acknowledgements

Funded by the Deutsche Forschungsgemeinschaft (DFG) Research Training Group 2386 (SW); EXC 2002/1 "Science of Intelligence" (KG); SFB 940 "Volition and Cognitive Control"; SFB-TRR 295 "Retuning dynamic motor network disorders using neuromodulation"; and supported by BMBF and Max Planck Society.

Correspondence

Please address correspondence to: Simon Weber, Bernstein Center for Computational Neuroscience, Charité – Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin (Email: sweber@bccn-berlin.de).

Abstract

It has been suggested that visual images are memorized across brief periods of time by vividly imagining them as if they still were there. Such visual imagery has been proposed as a key mechanism underlying working memory. However, the ability to evoke visual imagery varies substantially across individuals, raising the question whether people with weak or even absent visual imagery might use other, non-visual strategies for memorization. Here, we systematically investigated this question in two groups of participants, very strong and very weak imagers, that were asked to remember images across brief delay periods. We were able to reliably reconstruct the memorized stimuli from early visual cortex during the delay period. Importantly, the quality of reconstruction was equally accurate for both strong and weak imagers. The decodable information also closely reflected behavioral precision in both groups, suggesting it could potentially contribute to behavioral performance, even in the extreme case of completely aphantasic individuals. Our data thus suggest that strong and weak imagers do not differ in how early visual areas encode memory-related information and that this information might even be equally used in both groups.

Introduction

In recent years, visual imagery, the ability to generate pictorial mental representations in the absence of external visual stimulation (Kosslyn & Thompson, 2003; Pearson & Kosslyn, 2015), has received increasing attention as a potential mechanism for supporting visual working memory (Albers et al., 2013; Tong, 2013).

Both visual imagery and visual working memory have been linked to the encoding of information in early visual cortex (Dijkstra et al., 2019; Klein et al., 2004; Kosslyn & Thompson, 2003; Lee & Baker, 2016; Serences, 2016). The sensory recruitment hypothesis of visual working memory (D'Esposito & Postle, 2015; Sreenivasan et al., 2014) posits that visual information is maintained using selective activation patterns in early visual cortex. This matches with a common view of visual imagery, where early visual areas encode detailed, perception-like mental images via top-down connections from high-level regions (Dijkstra et al., 2017; Mechelli, 2004). Encoding of contents has been reported to be similar between perception and visual working memory (e.g. Ester et al., 2009; Harrison & Tong, 2009; Lee et al., 2013; Serences et al., 2009). This similarity has also been shown to hold between perception and imagery across multiple features, including orientations (Albers et al., 2013), objects (Lee et al., 2012; Ragni et al., 2020; Reddy et al., 2010), letters (Senden et al., 2019), and natural scenes (Naselaris et al., 2015). Furthermore, both visual working memory (Teng & Kravitz, 2019) and visual imagery (Pearson et al., 2008) can interfere with and bias perception of subsequent stimuli.

The similarities in cortical organization of imagery and visual working memory raise the question whether these two processes might be related or even share the same neural substrate. Indeed, it was directly shown for normal-viewing participants that visual working memory and imagery representations of orientations exhibit very similar neuronal activity patterns in early visual cortex (Albers et al., 2013) suggesting that visual working memory and visual imagery share a similar neural substrate (Tong, 2013). In this view, participants might

briefly memorize visual stimuli in working memory tasks by vividly imagining them across the delay period.

However, the ability to generate imagery as well as its vividness differ substantially across individuals (Kosslyn et al., 2001). Some people even report the complete absence of phenomenal imagery ("aphantasia"; Zeman et al., 2015; Zeman et al., 2010). Nonetheless, these differences do not appear to manifest themselves systematically in behavioral measures of memory. Rather, most studies indicate that behavioral performance in visual working memory tasks is comparable across imagery vividness levels, including the extreme case of aphantasic individuals (Jacobs et al., 2018; Zeman et al., 2015). However, differences have been reported. For example, working memory performance for strong imagers is disrupted by irrelevant visual input, while weak imagers show no such distraction effect (Keogh & Pearson, 2014), indicating the use of distinct memorization strategies. This is supported by comparing reports of strong and weak imagers. Strong imagers report to rely mostly on visual strategies when solving visual working memory tasks. In contrast, weak imagers tend to report using different cognitive strategies such as verbal or categorical associations (Bainbridge et al., 2021; Keogh et al., 2021; Logie et al., 2011). Thus, visual imagery might to be only one of several cognitive tools that can be used to solve visual working memory tasks. If this is true, then weak imagers could use different representational systems for maintaining stimulus features other than sensory recruitment in early visual cortex.

In line with this, the cognitive-strategies framework of working memory (Pearson & Keogh, 2019) postulates that the cognitive strategy used to solve a working memory task determines the format in which a stimulus is represented in the brain, and consequently influences how much information about the stimulus is present within a given cortical region. In the case of visual imagery, this could mean that individuals with high imagery vividness spontaneously recruit their early visual cortex to maintain detailed stimulus representations, while individuals with low imagery vividness employ alternative, non-visual strategies to solve the same

cognitive task. Together, this predicts that strong imagers should retain more information about a stimulus feature in their visual cortex activity than weak imagers.

Here, we directly test this hypothesis by assessing the influence of imagery vividness on the strength of visual working memory representations in visual cortex, using functional magnetic resonance imaging (fMRI). We recruited two groups of study participants, one with very high and one with very low imagery vividness scores as assessed by an established questionnaire (VVIQ, Marks, 1973; Fig. 1B; see Methods). In the main experiment, they performed a working memory task that involved memorizing a bright orientation stimulus across a brief delay (Fig. 1A). We used a brain-based decoder (periodic support vector regression; see Methods) to reconstruct these orientations from brain activity patterns in early visual cortex obtained during the memory delay period. If strong imagers indeed rely more on imagery signals in early visual cortex to maintain the stimulus across the delay, this could lead to two predictions: First, that sensory information should be represented more accurately in the early visual brain signals of strong as opposed to weak imagers; second, sensory information in early visual areas should also be more predictive of an individual's behavioral performance, especially in strong imagers.

Results

Behavioral results

Fig. 1C shows how accurately participants performed in the task. The figure plots the deviation between participants' judgements and the true orientations for each trial (grey bars), revealing that the responses were highly accurate. To assess this quantitatively, we fitted a computational model to the response distribution of each participant that yields estimates for behavioral precision and bias (von Mises mixture model; Fig. 1C, black line; see Methods for details). Across all participants, responses were precise (precision κ_1 = 5.673 ± 2.377), with a small but significant bias to respond anti-clockwise of the target (μ = -0.889° ± 1.635°; Fig. 1C, inset).

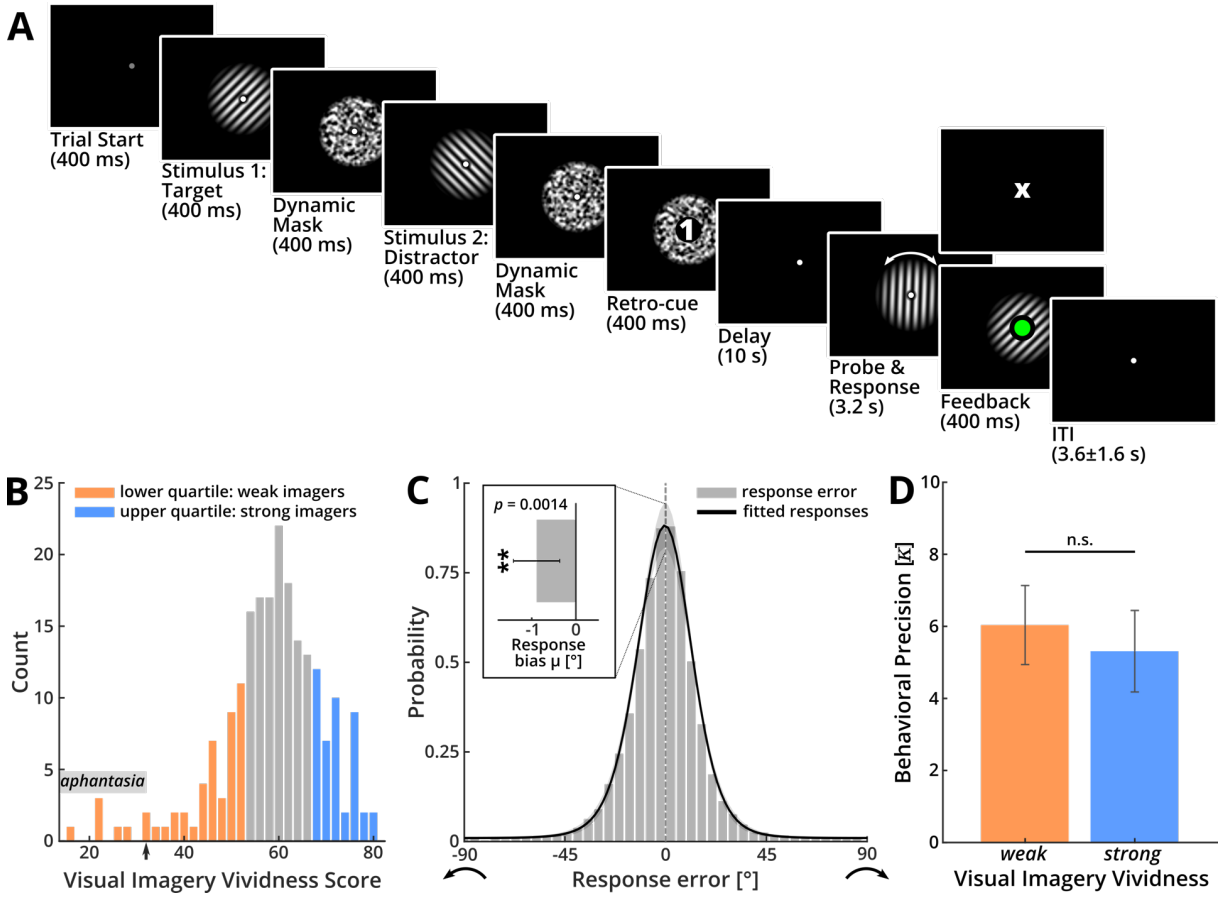

Fig. 1: Experimental task and behavioral results. (A) Sequence of events in one trial of the experiment. In each trial, participants were successively presented with two orientation stimuli, each followed by a dynamic noise mask. Orientations were drawn from a set of 40 discrete, equally spaced orientations between 0° and 180°. The stimuli were followed by a numeric retro-cue ("1" or "2"), indicating which one of them was to be used for the subsequent delayed-estimation task ("target"), and which could be dropped from memory ("distractor"). The orientation of the cued target grating had to be maintained for a 10-second delay. After the delay a probe grating appeared, which had to be adjusted using two buttons and then confirmed via an additional button press. Subsequently, visual feedback was given indicating whether a response was given in time (by turning the fixation point green, lower panel) or missed (by displaying a small "X" at the end of the response period if no response was given in time, upper panel). Cue and feedback are enlarged in this illustration for better visibility. **(B)** Distribution of the scores in an online visual imagery questionnaire (VVIQ, see Methods) that was used for recruitment. Subjects from the upper (blue) versus lower (orange) quartiles of the distribution were recruited for the strong and weak imagery vividness groups, respectively. The arrow on the x-axis points to the aphantasia cutoff. **(C)** Histogram of deviations between the reported and the true orientation of the target stimuli (grey bars) and a model fit of behavioral responses across all subjects (black line). For this, the responses were modeled using a von Mises mixture model (vMMM) for detections (responses to target orientations, assumed to follow a von Mises distribution with mean 0° plus bias μ and behavioral precision κ_1), swap errors (false responses to distractor orientations, following the same assumptions as detections) and guesses (assumed to follow a continuous uniform distribution between -90° and +90°). The model estimated individual probabilities for each of these three event classes (resulting in mixture coefficients, r_1, r_2 and r_3, respectively). The estimated parameters indicate that participants accurately performed the task: they correctly responded to the target direction in around 95 % of trials (r_1 = 0.947 ± 0.063). Across participants, responses were precise (κ_1 = 5.673 ± 2.377), with a small but significant bias to respond anti-clockwise of the target (inset; μ = -0.889 ± 1.635°; $t_{(39)}$ = -3.437, p = 0.0014, two-tailed; error bar: 95 % confidence interval).

See Fig. S1 for details on the other estimated parameters. **(D)** Behavioral precision (κ_1) for strong and weak imagers separately. Behavioral precision did not significantly differ between groups (error bars: 95 % confidence intervals).

Importantly, there were no significant differences between strong and weak imagers for behavioral precision (Fig. 1D; $t_{(38)}$ = -0.965, p = 0.341, two-tailed) or any other of the estimated behavioral parameters (Fig. S1). This indicates that the high individual differences in visual imagery were not associated with performance differences in the visual working memory task. We used a Bayesian analysis to assess the evidence for absence of a difference in behavioral precision between the weak and strong imagery groups. The Bayes factor indicated that the data were 2.2 times more likely under the null hypothesis (BF_{01} = 2.239) which provides weak evidence for the absence of an effect of imagery vividness on behavioral precision (Jeffreys, 1998).

Orientation reconstruction from fMRI data

We used a brain-based decoder to reconstruct orientation representations encoded in the patterns of signals in early visual cortex (V1-V3, see Fig. S2; see Methods). Across all subjects, we were able to reconstruct the true physical target orientation above chance-level for an extended period following delay onset (Fig. 2A, green line): At 5 s after delay onset, the accuracy rose to 12 % above chance, where it plateaued until 3 s after probe onset. Following probe onset, the accuracy increased steeply before falling back towards baseline. This later peak in reconstruction performance is likely to reflect the perceptual information of the adjustable probe grating after it had been rotated by the participants to report the target orientation. Reconstruction of the reported orientation yielded a very similar pattern of results (Fig. 2A, red line). This close resemblance was expected, given the close match between target and reported orientations (see Fig. 1C).

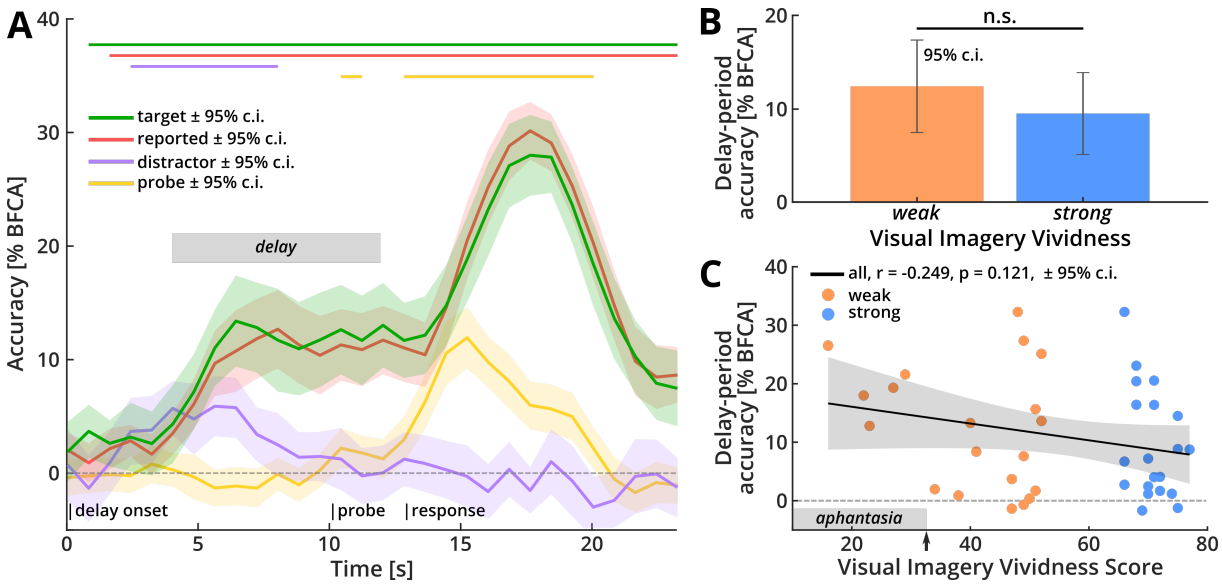

Fig. 2: Orientation reconstruction from early visual cortex. (A) Reconstruction performance for orientations based on brain signals from early visual areas V1-V3. The y-axis plots the accuracy (BFCA, see Methods), across time for target (green), reported (red), distractor (purple) and probe (yellow) orientations. The horizontal lines above the graph indicate time periods where this reconstruction was significantly above chance (permutation-based cluster-mass statistic, see Methods). The target orientation (green) could be reconstructed above chance-level throughout the delay and report periods (cluster-$p < 0.001$). Reconstruction of the reported orientation (red) followed a highly similar pattern (cluster-$p < 0.001$). The distractor orientation (purple) could only be reconstructed early in the trial (cluster-$p < 0.001$), before falling back to baseline. Reconstruction of the adjustable probe orientation (yellow) was only possible late in the trial (large cluster: cluster-$p < 0.001$; small cluster: cluster-$p = 0.015$), after it had been presented (shaded areas: 95 % confidence intervals). The gray box marks the preregistered delay-period time window used for subsequent analyses. **(B)** Target reconstruction performance for strong and weak imagers separately, pooled across the preregistered delay-period (gray bar in (A)). Delay-period decoding accuracy of strong imagers was not significantly increased compared to weak imagers ($t_{(38)} = 0.926$, $p = 0.82$, one-tailed; error bars: 95 % confidence intervals). **(C)** Detailed correlation between delay-period accuracy (BFCA) and visual imagery score. There was no significant correlation between the strength of delay-period representations and imagery vividness even when using the full graded imagery scores (shaded area: 95 % confidence interval). Neural information during the delay-period was significantly above chance-level even for aphantasic individuals with a visual imagery score below 32 (grey bar at x-axis; $t_{(4)} = 8.753$, $p < 0.001$, one-tailed, E.A.). The arrow on the x-axis points to the aphantasia cutoff.

We also conducted several checks to test for other predictions of our analysis. First, we reconstructed the orientation of the distractor, i.e., the task-irrelevant orientation stimulus that was not cued and could thus be forgotten after the retro-cue. As expected, information about this distractor orientation (Fig. 2A, purple line) was only present briefly at the beginning of the trial after which the accuracy returned to chance-level for the remainder of the trial. In line with previous work on the representation of task-irrelevant stimuli (e.g. Albers et al., 2013; Ester et al., 2013; Harrison & Tong, 2009), this transient early information presumably reflects the

perceptual signal following the presentation of the distractor early in the trial, delayed by the hemodynamic lag. Second, we reconstructed the initial random starting orientation of the adjustable probe grating (Fig. 2A, yellow line). As expected, this resulted in an informative time window late in the trial, after probe onset, likely reflecting the perceptual signal of the adjustable probe before it was rotated for the behavioral response. Together, this pattern of results indicates the presence of sustained, content-selective representations of the memorized stimuli during the delay-period, while task-irrelevant stimulus information was quickly dropped from memory. In an additional analysis we confirmed that the decodable information was not related to systematic eye-movements (Fig. S3).

Group differences in delay-period representations

Next, we proceeded to address the key question whether there was any indication that strong and weak imagers differed in their memory-related information in early visual cortex. Despite robust group-wise reconstruction performance, reconstruction accuracy of strong imagers was not significantly higher compared to weak imagers (Fig. 2B; $t_{(38)} = 0.926$, $p = 0.82$, one-tailed). This was confirmed by a post-hoc Bayesian *t*-test, which provided moderate evidence in favor of the null hypothesis over our original prediction ($BF_{01} = 5.55$).

We also performed additional tests to corroborate the effect. First, we re-assigned groups based on an additional post-scan questionnaire data (see Methods), which yielded similar results (Table S1). Second, we assessed the possibility that the effect of imagery vividness is more gradual in nature and thus might not be captured by the categorical group difference. To address this, we calculated the correlation between delay-period accuracies and graded imagery vividness scores. Again, the result was not significant (Fig. 2C, $r = -0.249$, $p = 0.121$), with strong evidence for the absence of a positive correlation ($BF_{01} = 12.209$). Note that delay-period accuracy was significantly greater than chance-level even for the five participants with a visual imagery score of below 32, which is generally considered the threshold for aphantasia (Zeman et al., 2015; marked with a grey bar on the x-axis of Fig. 2C; one-sample *t*-test: $t_{(4)}$ =

8.753, $p < 0.001$, one-tailed; E.A.). Taken together, these results suggest that imagery vividness, at least in the form of subjective questionnaire scores, does not affect the strength of delay-period representations of target orientations in early visual cortex.

Finally, we tested a further prediction that would be expected if strong imagers relied more on sensory information encoded in early visual cortex than weak imagers. In that case, there should be a tighter predictive link between behavioral performance and the encoding of information in early visual areas, especially for strong imagers. For this, we assessed whether there was more performance-predictive information in early visual areas of strong imagers. In this additional analysis (E.A.) we observed a strong correlation between delay-period accuracy and behavioral precision (Fig. 3A; $r = 0.739$, $p < 0.001$), which was the same across groups (Fig. 3B; strong: $r = 0.781$, $p < 0.001$; weak: $r = 0.695$, $p < 0.001$). Interestingly, half of the variance in delay-period accuracy could be explained by behavioral precision (R^2, all: 0.546; strong: 0.61; weak: 0.483). This strong effect suggests that the signals in early visual cortex could potentially play a direct role in maintaining the sensory stimulus across the memory delay (as suggested by the sensory recruitment hypothesis), and that this does not depend on whether a person is a strong or a weak imager.

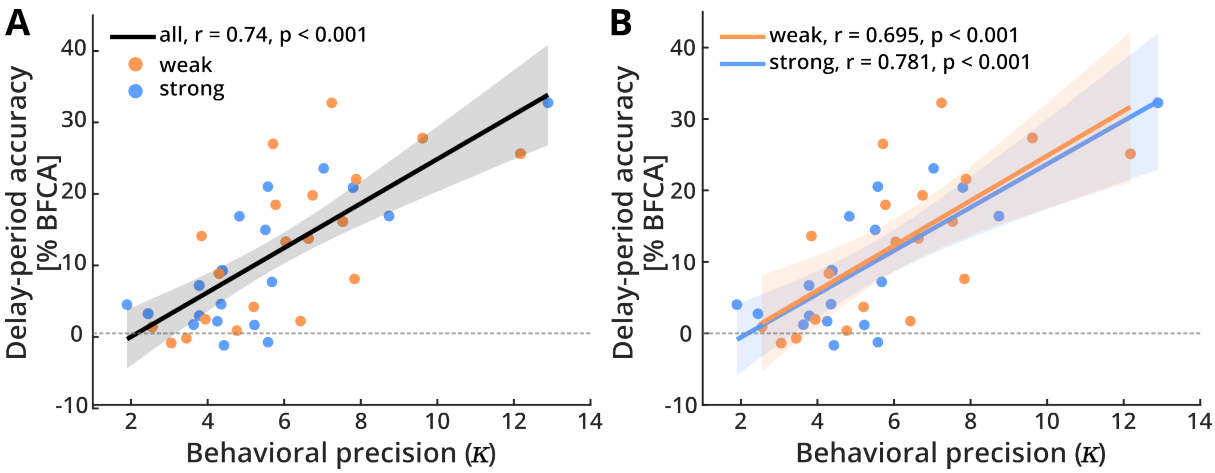

Fig. 3: Behavioral precision versus decodable neural information from early visual cortex. Correlation between the behavioral precision (kappa, κ_1) in the task and the accuracy of brain-based reconstruction. The strength of delay-period representations was highly predictable of behavioral precision, both **(A)** across all participants and **(B)** within strong and weak imagery vividness groups. Shaded areas indicate 95 % confidence intervals.

Discussion

In this study, we investigated to which extent an individual's visual imagery vividness affects the strength of memory representations in their visual cortex. Two experimental groups, strong and weak imagers, performed a visual working memory task, which involved memorizing images of oriented lines over a delay. In both groups we found that early visual cortex contained robust information about the remembered orientations across the entire delay period. Importantly, testing our primary research question, the level of this information did not differ between strong and weak imagery groups. There was also no apparent dependency of visual cortex representations on any other subjective measures of encoding strategy (see Table S1), suggesting that remembered orientations were encoded equally strongly in the visual areas irrespective of imagery vividness. Crucially, even the five participants with a VVIQ score of below 32, which is generally considered the threshold for complete absence of phenomenal imagery ("aphantasia", Zeman et al., 2015), showed comparable visual neural information to the strong imagers (see Fig. 2C).

We also found no evidence that these signals in early visual cortex might be *utilized* differently across imagery levels. We observed a strong correlation between encoded information and individual behavioral precision, in line with a close link between neural information and task performance (see below). However, the overall strength of this effect was also indistinguishable between imagery groups. We thus find no evidence for differences between strong and weak imagers, neither in the encoding of sensory information nor in our indicator of how this information informs behavior. Please note that these null effects were based on preregistered analyses and are supported by additional Bayesian analyses.

Our observation of a close link between sensory information in the delay period and behavioral working memory performance is in line with several previous studies (e.g. Bettencourt & Xu, 2016; Ester et al., 2013; Hallenbeck et al., 2021; Harrison & Tong, 2009; Iamshchinina et al., 2021). Based on our highly sensitive method to reconstruct continuous stimulus features from

voxel patterns, the neural information explained more than half of the between-subject variance in behavioral performance (see Methods for more details), which further corroborates the link between information encoded in early visual cortex and memorization of visual information across brief delays. Additionally, we found that sensory information was retained only for the cued and thus task-relevant stimulus but was not present for the uncued image. These results are in line with sensory recruitment accounts of working memory (D'Esposito & Postle, 2015), or more generally with a multi-level representation of sensory information across delays (Christophel et al., 2017), according to which sensory cortices contribute to working memory maintenance by facilitating detailed, perception-like representations of stimulus features that are specifically relevant for the current task. This task-dependent retention of information in early visual cortex could point towards some form of *active maintenance* throughout the delay after offset of the stimulus. This could be achieved by neural mechanisms such as recurrent processing within early visual cortex (Lamme & Roelfsema, 2000) or by feedback from higher regions (Gazzaley & Nobre, 2012) and could include short-term synaptic plasticity (Mongillo et al., 2008; Rose et al., 2016). However, these mechanisms are below the resolution of neuroimaging and thus remain hypothetical for our study.

Our key finding is the observation of task-dependent maintenance in the visual cortices of weak imagers and even in aphantasic individuals, who report to be largely unable to evoke visual images and tend to use non-visual (such as propositional or phonological) strategies to encode task-relevant stimuli (Bainbridge et al., 2021; Keogh et al., 2021). Note, that our study focused on encoding of working memory representations and did not directly assess the encoding of mental imagery contents. We cannot explicitly assess whether weak and strong imagers used different *strategies* to maintain the stimulus information during the task because the two groups did not receive different encoding instructions. Our results nevertheless clearly indicate that any potential differences in encoding strategies, as well as in experienced imagery vividness, did not affect encoding of sensory content during the delay period in early visual areas. This goes directly against our key prediction from the cognitive-strategies framework of working

memory (Pearson & Keogh, 2019), according to which strong imagers should retain higher levels of stimulus information in their early visual cortices during working memory, compared to weak imagers. Our results therefore call into question the role of experienced imagery vividness in the modulation of early visual cortex recruitment during working memory.

An interesting and related question is whether our data bear any information on whether early visual cortex is essential for visual imagery, regarding which there has been some debate. On the one hand it has been argued that visual imagery is based on higher-level representations beyond early visual cortex (Bartolomeo, 2008). For instance, vivid phenomenal imagery can be preserved in cortically blind patients after strokes to occipital areas (Bartolomeo et al., 1998; Chatterjee & Southwood, 1995; de Gelder et al., 2015), indicating that early visual cortex is not essential for visual imagery. Similarly, brain damage can selectively affect visual imagery but leave visual perception largely preserved. This has been reported after selective lesions in temporal regions (Moro et al., 2008; Thorudottir et al., 2020), and has been taken as evidence that visual imagery depends on a temporal network (Spagna et al., 2021). Together, this would suggest a functional dissociation of early visual cortex and visual imagery (Bartolomeo et al., 2020). On this basis, one could speculate that the weak imagers in our case might have had a deficit in this temporal imagery-related network whereas memory performance is based on sensory information that is largely intact. Early visual information would thus be available to solve the working memory task but would not necessarily lead to the experience of imagery (please note that the encoding of imagery contents was not the goal of our study). On the other hand, this is at odds with current interpretations of behavioral, neuroimaging and brain-stimulation work which suggest a close link between signals in early visual areas and imagery (Albers et al., 2013; Keogh et al., 2020; Pearson, 2019). One potential reason for this discrepancy may lie in the different ways in which imagery vividness is quantified across the literature (Pearson, 2020): Much of the brain-lesion work has relied on subjective vividness measures which tend to assess imagery of high-level features such as object or complex scenery, whereas more objective measures, mostly employed in behavioral work, target

imagery of low-level sensory features such as orientations (Pearson et al., 2008). However, previous studies suggest that subjective (high-level) and objective (low-level) measures of imagery vividness tend to correlate (Bergmann et al., 2016; Dijkstra, Bosch, et al., 2017; Pearson et al., 2011), and people generally provide good metacognitive judgments about their own imagery abilities (Pearson et al., 2011; Rademaker & Pearson, 2012).

In conclusion, we show that the active maintenance of stimulus-related information in early visual areas was present also in participants who report a near-absence of visual imagery. The encoding of sensory information and its link to performance was strong and indistinguishable across different levels of imagery. This provides further evidence for the view that differences in subjectively experienced imagery do not influence recruitment of early visual cortex for working memory. Instead, it appears that working memory representations in visual cortex are maintained irrespective of whether a person is able to engage in vivid imagery or not.

Methods

Preregistration

The main analysis workflow of this study (including custom preprocessing steps, parameter choices, ROIs and newly implemented statistical models) was preregistered at https://osf.io/34y9z. The preregistration was submitted after data acquisition, but prior to data processing and analysis. All preregistered analysis procedures were developed and/or optimized on a separate fMRI dataset from a related study (Barbieri et al., 2023). Please note that we did not change any of the preregistered workflows. However, we did perform additional analyses and performed more extended statistical testing (e.g., Bayesian and permutation-based tests) whenever it proved necessary to the quality of the study. All of these additional analyses are indicated as E.A. (“extended analysis”) in this text.

Recruitment

Two groups of study participants were preselected for the study using an online version of the Vividness of Visual Imagery Questionnaire (VVIQ; Marks, 1973). The questionnaire was implemented and hosted on the online survey platform SoSci Survey (www.soscisurvey.de) and local respondents were recruited via in-house mailing lists for experimental studies, study participant databases and Facebook. Respondents gave informed consent prior to being directed to the questionnaire and again before providing an email address for recruitment at the end of the questionnaire.

We received a total of 263 online responses, 210 of which fulfilled the physiological, medical and demographic criteria for participation in the MRI study. Respondents whose VVIQ scores fell either into the upper or lower quartiles of the response distribution were assigned to the strong and weak imagery groups, respectively, and contacted for participation in the fMRI experiment (Fig. 1B). From these groups we recruited a total of 42 fMRI participants. All participants were healthy, right-handed individuals between 18 and 40 years with no history of neurological or psychiatric disorders. One participant dropped out of the study before completing all scanning sessions. The data of a second participant had to be discarded due to technical issues with the MRI scanner. Therefore, we collected complete datasets of 40 participants (female: 23, age: 28.05 ± 6.064 years), 20 each per experimental group (average VVIQ score; strong: 70.7 ± 3.262; weak: 40.75 ± 11.571).

Participants gave written informed consent prior to the fMRI experiment. They received monetary compensation of 10€/h for the fMRI sessions and a bonus of 10€ for completion of both scanning sessions. Following April 19, 2021, participants were required to present a negative SARS-CoV-19 rapid test result (not older than 24 hours) before entering the MRI facility. To compensate for the additional effort, we paid an additional 20€ for each SARS-CoV-19 rapid test. The study was approved by the ethics committee of the Humboldt-Universität zu Berlin and conducted according to the principles of the Declaration of Helsinki (World Medical Association, 2013).

Stimuli

The experiment was implemented using MATLAB 2018b (The MathWorks, Inc.) and Psychtoolbox 3 (Brainard, 1997; Kleiner et al., 2007). All stimuli were presented on black background, to avoid residual luminance interfering with potential visual imagery during the delay period (Keogh & Pearson, 2014). For stimulation, we used circular high contrast sine-wave Gabor patches with phase 0, contrast 0.8 and a spatial frequency of 0.02 cycles per pixel. Stimuli were presented inside a circular aperture with an inner diameter of 0.71 dva and an outer diameter of 8.47 dva. A white fixation dot of 0.18 dva was placed at the center of the inner aperture (Fig. 1A).

The set of target orientations comprised 40 discrete, equally spaced orientations separated by 180°/40 = 4.5°. To avoid the exact cardinal directions (0°, 45°, 90°, 135°), the orientations were slightly shifted by 1.125°, resulting in a set of orientations between 1.125° and 176.625°. Another set of 40 gratings, which served as distractors, was created by shifting the target orientations by 4.5°/2 = 2.25°, yielding orientation stimuli between 3.375° and 178.875°. This ensured that (i) target and distractor orientations were never exactly the same and (ii) both sets of orientations avoided the exact cardinal directions. Since we presented 40 trials in each run (see below), each target and distractor orientation was shown once during each run, in randomized order. Accordingly, target and distractor orientations were counterbalanced across runs. The starting orientation of the probe grating was randomly selected from a uniform distribution between 0° and 180° on each trial.

To avoid afterimages, we used a custom dynamic noise mask (Fig. 1A). For each presentation of the mask, we initialized a 42-by-42 array of an equal number of black and white squares. Each time the screen was refreshed (refresh rate: 60 Hz), the array was scrambled along the rows and columns and smoothed by convolving it with a 2 x 2 box blur kernel. This created a highly dynamic noise mask that reliably suppressed afterimages of the high-contrast gratings. Masks were presented inside the same circular aperture as the stimuli.

fMRI task

The visual stimuli were presented on an MRI-compatible monitor (dimensions: 52 x 39 cm, resolution: 1024 x 768 px), positioned at the far end of the scanner bore, and viewed via an eye-tracking compatible mirror mounted on top of the head-coil. The distance between the eyes and the center of the monitor was 158 cm.

Each trial of the experiment started with the presentation of a central fixation dot which remained visible throughout the entire trial (Fig. 1A). Participants were instructed to fixate the dot at all times. After 0.4 s, participants were sequentially presented with two gratings (see above), one serving as the target and the other as the distractor. Each grating was shown for 0.4 s, followed by 0.4 s of a dynamic, high-contrast noise mask to avoid after-images. After the second mask a numerical retro-cue (0.4 s) was presented at the location of the fixation dot, indicating to the participants to remember the orientation of either the first ("1") or second ("2") grating during the subsequent delay period. The delay period lasted for 10 s, during which only the fixation dot remained visible on the screen. After the delay, a probe grating with random starting orientation appeared for 3.2 s. Participants were asked to adjust the orientation of the probe grating in a way that it corresponded to the remembered (target) orientation, using two buttons with the index and middle fingers of their right hand. After adjustment, participants had to confirm their response by pressing a button with the index finger of their left hand. If the response was completed within the time-window of 3.2 s, the fixation dot turned green for the remainder of the response period as visual feedback. If participants failed to provide a response in time, a small "X" was presented at the location of the fixation dot for 0.4 s. Trials were separated by a variable inter-trial interval (ITI) of 3.6 ± 1.6 s. Participants completed 40 trials per run and a total of 8 runs, equally split across 2 fMRI sessions on separate days, resulting in 320 trials per participant.

MRI data acquisition

MRI data were collected with a 3-Tesla Siemens Prisma MRI scanner (Siemens, Erlangen, Germany) using a 64-channel head coil. At the beginning of each session, we recorded a high-

resolution T1-weighted MPRAGE structural image (208 sagittal slices, TR = 2400 ms, TE = 2.22 ms, flip angle = 8°, voxel size = 0.8 mm^2 isotropic, FOV = 256 mm). On each of the two days, this was followed by four experimental runs, for each of which we recorded a series of 965 T2-weighted functional images using a multi-band accelerated EPI sequence with a multiband factor of 8 (TR = 800 ms, TE = 37 ms, flip angle = 52°, voxel size = 2 mm^2 isotropic, 72 slices, 1.9 mm inter-slice gap), resulting in a duration of 12:52 min per run. The first four TR of each sequence were discarded.

Eye-tracking

We used an EyeLink 1000 Plus (SR-Research) eye-tracker to record gaze position and pupil size of the dominant eye of each participant during the experimental runs. The tracker was positioned at the far end of the scanner bore (eye-lens-distance: 85 cm) on a long-distance mount and was calibrated once at the beginning of each session. Due to technical difficulties, we were only able to record eye-tracking data of 26 participants (13 per experimental group).

Post-experiment questionnaires

After the second session, participants completed three questionnaires: (i) a strategy questionnaire, asking (on a 5-point scale) for the degree to which they had used specific mnemonic strategies, such as visual, verbal, spatial or associative (e.g. imagining a clock face), to remember the target orientations and complete the task; (ii) the Vividness of Visual Imagery Questionnaire (VVIQ, as a post-experimental reference); and (iii) the Object-Spatial Imagery Questionnaire (OSIQ; Blajenkova et al., 2006), a 30-item questionnaire probing the strength of visual and spatial imagery.

Behavioral data analysis

Behavioral responses were modeled by fitting a von Mises mixture model (vMMM) to the distribution of behavioral response errors (Töpfer et al., 2022; original code available at https://github.com/JoramSoch/RDK_vMMM). The model is inspired by previous work on

modelling detections from working memory with similarly continuous features (Zhang & Luck, 2008). In our case we assume that on every trial participants either *detect* the target (responses to target orientations, assumed to follow a von Mises distribution with mean 0° plus bias μ and precision κ), make a *swap error* (responses to distractor orientations, following the same assumptions as detections) or *guess* (assumed to follow a continuous uniform distribution between -90° and +90°). Each of these three potential trial-wise outcomes (detections, swaps and guesses) has an associated probability distribution indicating how probable each potential response angle is, given the orientation of the stimulus (i.e., target and distractor). The overall response distribution is considered a linear combination of these three individual event probability distributions with associated probabilities as mixture coefficients r_1, r_2 and r_3.

According to this approach, the probability of observing a specific response evaluates to

$$P\left(\theta_r \middle| \theta_t, \theta_d, \vec{r}, \mu, \vec{\kappa}\right) = P(\text{'detection'}) \cdot P(\theta_r|\text{'detection'}) + P(\text{'swap'}) \cdot P(\theta_r|\text{'swap'}) + P(\text{'guess'}) \cdot P(\theta_r|\text{'guess'})$$
$$= r_1 \frac{\exp\left[\kappa_1 \cos((\theta_t - \theta_r)_{\text{circ}} \frac{\pi}{90} - \mu)\right]}{2\pi I_0(\kappa_1)} + r_2 \frac{\exp\left[\kappa_2 \cos((\theta_d - \theta_r)_{\text{circ}} \frac{\pi}{90} - \mu)\right]}{2\pi I_0(\kappa_2)} + r_3 \frac{1}{2\pi}$$

where θ_r is the reported orientation in degrees; θ_t, θ_d and are the target and distractor orientations in degrees, respectively; $\vec{r}$ is a vector containing r_1, r_2 and r_3, the event probabilities for the three model components (detections, swap errors and guesses); $\vec{\kappa}$ is a vector containing κ_1 and κ_2, the precisions for detections and swap errors, respectively; μ is the response bias and $I_0(\kappa_i)$ is the modified Bessel function of order 0. As κ_1 reflects the width of the response distribution for target detections, we report this parameter as our key measure for behavioral precision.

fMRI data analysis: preprocessing

Processing and analysis of fMRI data was performed in MATLAB, using SPM12, The Decoding Toolbox (Hebart, Görgen et al., 2015) and custom scripts (see below). MR images were converted into NIfTI format for further processing. Before the analysis, BOLD images were

spatially realigned and resliced. The T1 image of each session was coregistered to the first image of the respective BOLD series. We then calculated normalization parameters to the Montreal Neurological Institute (MNI) standard space. These were used to project probabilistic maps of our regions of interest (ROIs) into the native space of each individual participant to guide voxel selection during the reconstruction analysis (see below). Following realignment, the time series of each voxel's raw data were temporally detrended, to remove slow signal drifts that accumulate across a given run. This was implemented using cubic spline interpolation (modifying an algorithm proposed by Tanabe et al., 2002). The time series of voxel data for a given run was separated into 40/2 = 20 segments of equal size. The data from each segment was averaged to create query points (nodes), which were then used for cubic spline interpolation, creating a smooth function modeling the slow signal drifts in the voxel data across the run. The number of nodes was specifically set to half the number of trials per run, to avoid the modeling (and thereby, removal) of within-trial effects. The drift-estimate was then subtracted from the voxel data. This procedure was repeated for every voxel and every run. After detrending, we applied temporal smoothing to the data by running a moving average of width 3 TR across the data of each run.

To increase the signal-to-noise ratio for samples from trials with neighboring stimulus orientations, we developed a method that we refer to as "feature-space smoothing". Feature-space smoothing accounts for the assumption that, in a feature-continuous stimulus space, samples that lie closely together in feature space (e.g., neighboring orientations) should produce a similar neural response and therefore a similar voxel signal. By reducing the contribution of noise to the measurements of neighboring samples, it should be possible to increase the amount of information represented in the voxel signal across the feature space. We addressed this issue by using a gaussian smoothing kernel to compute a weighted average of the voxel signal corresponding to a given orientation and its neighbors (Fig. S4). This means that samples close to a given orientation in feature space contribute more to the resulting average than those further away. The number (or distance) of samples included in the average is determined by the width (full width at half maximum, FWHM) of the smoothing kernel. Please

note that we confirmed through simulations that feature-space smoothing can substantially increase signal-to-noise ratio and thereby reconstruction accuracies without producing spurious above-chance accuracies in the case of null data (Fig. S5). In this study, we used nested cross-validation across subjects to determine the optimal kernel width for each participant (see below). Please note that all these approaches for temporal detrending and feature-space smoothing were developed and optimized on a separate dataset (from a related study) and both were pre-registered and checked for artifacts or spurious effects.

Anatomical and functional definition of the early visual cortex ROI

As our goal was to determine the strength of working memory representations in visual sensory stores depending on visual imagery vividness, we restricted our analysis to visually driven voxels in early retinotopic visual cortex (V1, V2, V3). In a first step we combined the probabilistic anatomical maps of V1, V2 and V3 (Wang et al., 2015) creating a combined map in standard space, collapsing across left and right hemispheres (Fig. S2). We then transformed this map into the native space of each participant, by applying the inverse normalization parameters estimated during preprocessing. The individual maps were then thresholded at 0.1, to exclude voxels that had a less than 10 % probability of being part of a given area, and binarized. This resulted in an average ROI size of 5938.6 ± 858.45 voxels. In a second step we identified visually driven voxels within that ROI. For this, we estimated a GLM with target onsets as event regressors of interest (plus additional event regressors of no interest for distractor, cue, delay and probe). All regressors were convolved with a canonical hemodynamic response function. We also added the 6 head motion time series. The purpose of this model was to calculate the target versus baseline contrast, in order to determine voxels with significant activation in response to the target, irrespective of orientation. The resulting statistical parametric maps were then used in combination with the individual anatomical ROIs for voxel selection in the multivariate reconstruction analysis. For this, we selected the voxels rank-ordered by their respective *t*-score (from the unspecific target versus baseline contrast)

within the anatomical ROI for each individual. The cutoff yielding the exact number of voxels was determined via nested cross-validation across subjects (see below).

Orientation reconstruction from fMRI data

The aim of our reconstruction analysis was to predict the angle of the orientation stimulus from the multivariate signal of the preprocessed raw data in the early visual cortex ROI. Note that the space of orientations is circular between 0° and 180°. To account for this, we implemented periodic support vector regression (pSVR), a periodic extension of the SVR (Drucker et al., 1996). First, we projected the angular labels into a periodic space by calculating two sinusoids in the range [0°, 180°). Both functions had an amplitude of 1 and a period of 180°, so that one period spanned the entire label space. One function was shifted by 45°, so that the combination of both periodic functions coded for the linear label scale (Fig. S6). This is the 180°-equivalent to the way sine and cosine functions between 0° and 360° code for the angles on a unit circle. Next, we individually predicted each set of labels from the multivariate voxel pattern using the LIBSVM (Chang & Lin, 2011) implementation of SVR with a non-linear radial basis function (RBF) kernel, via a leave-one-run-out cross-validation. Before prediction, the voxel signals in the training data were rescaled to the range [0, 1]. The scaling parameters were then applied to the test data ("across-scaling", see Hebart et al., 2015).

After prediction of both sets of periodic labels $(\widehat{x_\iota}, \widehat{y_\iota})$ we computed the reconstructed angular orientation $\widehat{\theta_\iota}$ using the four-quadrant inverse tangent:

$$\widehat{\theta_\iota} = \mathrm{atan2}(\widehat{x_\iota}, \widehat{y_\iota})$$

The analysis was repeated for the 30 TRs (24 s) following delay-onset, for each TR individually. This allowed for a time-resolved estimation of how orientations were represented in the visual cortex across the entire trial.

Reconstruction performance evaluation

To evaluate the accuracy of the orientation reconstruction, we computed the feature-continuous accuracy (FCA). FCA is a rescaling of the absolute angular deviation (between predicted and true label) into the range 0-100 % and can be calculated, for the case of stimuli that are 180°-periodic, as (Pilly & Seitz, 2009)

$$\mathrm{FCA}(\theta_i, \widehat{\theta_i}) = \frac{90 - \left|(\theta_i - \widehat{\theta_i})_{\mathrm{circ}}\right|}{90} * 100$$

where θ_i is the true orientation in the i-th trial and $\widehat{\theta_i}$ is the associated reconstructed orientation. This trial-wise measure of reconstruction performance can be easily interpreted as a feature-continuous analogue to the accuracy measure of more conventional classification approaches: a value of 100 % means that there is no deviation between true and reconstructed orientations, i.e. perfect reconstruction; 50 % means deviation of 45°, which for circular orientation data is equivalent to guessing and can be considered as the chance-level; and 0 % means that reconstructed and true orientations are exactly orthogonal. FCA can be averaged to quantify reconstruction accuracy across trials.

For behavioral responses, the orientation labels may not be uniformly distributed across the orientation space, but clustered around, for example, cardinal axes. In a reconstruction setting, this would be analogous to a classification case with unequal (or unbalanced) numbers of classes, where the predictive model can exploit the uneven distribution of classes to simply predict the more frequent class more often. To account for this potential source of bias, we calculated a balanced FCA (BFCA). BFCA is an extension of the concept of balanced accuracy (Brodersen et al., 2010) for continuous variables. It is calculated by computing the integral of the trial-wise FCA from 0° to 180° (i.e. the orientation-space), using trapezoidal numerical integration across the sorted true and reconstructed orientations (Barbieri et al., 2023):

$$\mathrm{BFCA} = \frac{1}{180} \int_{0}^{180} \mathrm{FCA}(\theta, \hat{\theta})\, \mathrm{d}\theta$$

The process of integration assigns lower weights to the FCA values in the well-populated parts of the label-distribution and higher weights to the less populated parts. Thus, BFCA is a non-trial-wise measure of reconstruction performance, which accounts for the potential bias in FCA caused by non-uniformly distributed labels. We report BFCA as our key measure for reconstruction accuracy. Note that this approach has been previously tested to exclude the possibility of artifactual results.

Parameter optimization

As mentioned above, we used an across-subjects nested cross-validation to determine the optimal values of two parameters for each participant individually: (i) the width of the gaussian kernel used for feature-space smoothing, and (ii) the number of voxels entered into the analysis. For (i), we chose FWHM values between 0° (i.e., no smoothing) and 90°, in steps of 10°. Thus, we had a set of 10 possible kernel widths for smoothing. For (ii), we chose voxel counts between 250 and 2500, in steps of 250. This resulted in a set of 10 possible voxel counts. To select the specific voxels entered into the analysis, we first masked the individual target-versus-baseline *t*-maps with the warped anatomical ROIs (see above) and then selected the n voxels with the highest *t*-scores within those ROIs, with n representing a number from the set of possible voxel counts. The reconstruction analysis was then run for every combination of FWHM values and voxel counts.

After reconstruction, we determined the optimal parameters for each subject in the following way: First, we calculated the mean BFCA across all remaining subjects for every parameter combination, resulting in one value per combination and time point. Second, we averaged across the preregistered delay-period TRs (TRs 6-15 following delay onset), as we were specifically interested in potential group differences during this time window. This yielded one BFCA value per parameter combination, specifically for the entire delay period. The parameter

combination that yielded the highest BFCA was then assigned to the left-out subject. This was repeated for every subject and resulted in an average FWHM value of 74.5 ± 9.04 and an average voxel count of 1750 ± 211.83.

Statistical Testing

As we were specifically interested in potential group differences during the delay-period, statistical testing for differences between the strong and weak imagery groups was based on the time points in the trial which most likely only reflect delay period activity. Since the canonical hemodynamic response has a buildup of ~5 seconds, we considered the TRs 6-15 in the 30 TR timeframe that we analyzed, corresponding to a time window of 4 s after delay onset to 2 s after probe onset. This preregistered time window should avoid the leaking of stimulus- or probe-representations into the delay-period analysis.

We used two-tailed two-sample *t*-tests to test for potential differences in the reconstruction scores between the experimental groups. Further, we calculated Pearson's *r* to assess the correlation between outcome variables (E.A.).

Cluster-based permutation approach (E.A.) We were interested at which time points during the trial we could detect significant above-chance reconstruction accuracy. To account for the multiple-comparisons (Groppe et al., 2011) and autocorrelation (Purdon & Weisskoff, 1998) issues that arise for such time-resolved analyses, we adopted a non-parametric cluster-based permutation approach (Bullmore et al., 1999; Groppe et al., 2011; Maris & Oostenveld, 2007). This procedure was performed after the parameter optimization described above, to restrict the time-consuming permutation analysis to one set of parameters per subject. We repeated this approach separately for each reconstructed label type: target, distractor, probe and reported orientation.

Bayesian tests (E.A.). As our results indicated no significant differences between our two groups, we used Bayesian hypothesis tests to assess the evidence for this absence. Bayesian

hypothesis tests are used to describe the probability of observing the measured data under the null and alternative hypothesis, respectively (Keysers et al., 2020). This likelihood is quantified using the Bayes factor (BF), a continuous measure of evidence for either hypothesis. Specifically, we used two Bayesian hypothesis tests to assess the evidence for absence of effects: First, in the case of non-significant group-comparisons, we performed follow-up Bayesian independent *t*-tests, using a Cauchy distribution with scale parameter $r = 0.707$ as the prior distribution (Morey & Rouder, 2011). Second, in the case of non-significant correlations, we performed Bayesian correlation with a stretched beta prior of width $\kappa = 1$. All Bayesian hypothesis tests were performed in the open-source software JASP (Love et al., 2019).

Orientation reconstruction from eye-tracking data

Participants were instructed to maintain fixation at all times during the experiment. It is at least theoretically conceivable that participants might have used an eye-movement-based strategy to remember target orientations. Eye-movements have also been shown to modulate visual responses in the brain (Merriam et al., 2013). To account for these potentially confounding factors, we investigated whether the gaze position across the trial held information about the target orientation. For this, we subjected the recorded x and y ordinates of 26 participants (for which complete sets of eye-tracking data were available) to the same reconstruction analysis as the fMRI data.

Preprocessing of eye-tracking data was performed in MATLAB using functions from the Fieldtrip toolbox (Oostenveld et al., 2011), code adapted from prior work (Urai et al., 2017) and in-house code. Blinks were linearly interpolated and bandpass filtered between 5 Hz (high-pass) and 100 Hz (low-pass). For each trial, we extracted 15 s worth of data following the onset of the first grating. The data from each run was detrended using the same cubic spline interpolation as described above (see Preprocessing of fMRI data). We then downsampled the data by a factor of 10, resulting in 1500 time points per trial.

After preprocessing, we entered the data into the same pSVR reconstruction analysis as the fMRI data, using the x and y ordinates of the gaze position as input instead of voxel signal, and evaluated the reconstruction by calculating the BFCA. As with the fMRI data, we tested for clusters of above-chance time points using the cluster-based *t*-mass permutation approach described above.

References

Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C., & de Lange, F. P. (2013). Shared Representations for Working Memory and Mental Imagery in Early Visual Cortex. *Current Biology*, *23*(15), 1427–1431. https://doi.org/10.1016/j.cub.2013.05.065

Bainbridge, W. A., Pounder, Z., Eardley, A. F., & Baker, C. I. (2021). Quantifying aphantasia through drawing: Those without visual imagery show deficits in object but not spatial memory. *Cortex*, *135*, 159–172. https://doi.org/10.1016/j.cortex.2020.11.014

Barbieri, R., Töpfer, F. M., Soch, J., Bogler, C., Sprekeler, H., & Haynes, J.-D. (2023). *Encoding of continuous perceptual choices in human early visual cortex* [Preprint]. bioRxiv. https://doi.org/10.1101/2023.02.10.527876

Bartolomeo, P. (2008). The neural correlates of visual mental imagery: An ongoing debate. *Cortex*, *44*(2), 107–108. https://doi.org/10.1016/j.cortex.2006.07.001

Bartolomeo, P., Bachoud-Lévi, A.-C., De Gelder, B., Denes, G., Dalla Barba, G., Brugières, P., & Degos, J.-D. (1998). Multiple-domain dissociation between impaired visual perception and preserved mental imagery in a patient with bilateral extrastriate lesions. *Neuropsychologia*, *36*(3), 239–249. https://doi.org/10.1016/S0028-3932(97)00103-6

Bartolomeo, P., Hajhajate, D., Liu, J., & Spagna, A. (2020). Assessing the causal role of early visual areas in visual mental imagery. *Nature Reviews Neuroscience*, *21*, 2. https://doi.org/10.1038/s41583-020-0348-5

Bergmann, J., Genç, E., Kohler, A., Singer, W., & Pearson, J. (2016). Smaller Primary Visual Cortex Is Associated with Stronger, but Less Precise Mental Imagery. *Cerebral Cortex*, *26*(9), 3838–3850. https://doi.org/10.1093/cercor/bhv186

Bettencourt, K. C., & Xu, Y. (2016). Decoding the content of visual short-term memory under distraction in occipital and parietal areas. *Nature Neuroscience*, *19*(1), 150–157. https://doi.org/10.1038/nn.4174

Blajenkova, O., Kozhevnikov, M., & Motes, M. A. (2006). Object-spatial imagery: A new self-report imagery questionnaire. *Applied Cognitive Psychology*, *20*(2), 239–263. https://doi.org/10.1002/acp.1182

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Brodersen, K. H., Ong, C. S., Stephan, K. E., & Buhmann, J. M. (2010). The Balanced Accuracy and Its Posterior Distribution. *2010 20th International Conference on Pattern Recognition*, 3121–3124. https://doi.org/10.1109/ICPR.2010.764

Bullmore, E. T., Suckling, J., Overmeyer, S., Rabe-Hesketh, S., Taylor, E., & Brammer, M. J. (1999). Global, voxel, and cluster tests, by theory and permutation, for a difference between two groups of structural MR images of the brain. *IEEE Transactions on Medical Imaging*, *18*(1), 32–42. https://doi.org/10.1109/42.750253

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, *2*(3), 1–27. https://doi.org/10.1145/1961189.1961199

Chatterjee, A., & Southwood, M. H. (1995). Cortical blindness and visual imagery. *Neurology*, *45*(12), 2189–2195. https://doi.org/10.1212/WNL.45.12.2189

Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R., & Haynes, J.-D. (2017). The Distributed Nature of Working Memory. *Trends in Cognitive Sciences*, *21*(2), 111–124. https://doi.org/10.1016/j.tics.2016.12.007

de Gelder, B., Tamietto, M., Pegna, A. J., & Van den Stock, J. (2015). Visual imagery influences brain responses to visual stimulation in bilateral cortical blindness. *Cortex*, *72*, 15–26. https://doi.org/10.1016/j.cortex.2014.11.009

D'Esposito, M., & Postle, B. R. (2015). The Cognitive Neuroscience of Working Memory. *Annual Review of Psychology*, *66*(1), 115–142. https://doi.org/10.1146/annurev-psych-010814-015031

Dijkstra, N., Bosch, S. E., & van Gerven, M. A. J. (2017). Vividness of Visual Imagery Depends on the Neural Overlap with Perception in Visual Areas. *The Journal of Neuroscience*, *37*(5), 1367–1373. https://doi.org/10.1523/JNEUROSCI.3022-16.2016

Dijkstra, N., Bosch, S. E., & van Gerven, M. A. J. (2019). Shared Neural Mechanisms of Visual Perception and Imagery. *Trends in Cognitive Sciences*, *23*(5), 423–434. https://doi.org/10.1016/j.tics.2019.02.004

Dijkstra, N., Zeidman, P., Ondobaka, S., van Gerven, M. A. J., & Friston, K. (2017). Distinct Top-down and Bottom-up Brain Connectivity During Visual Perception and Imagery. *Scientific Reports*, *7*(1), 5677. https://doi.org/10.1038/s41598-017-05888-8

Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A. J., & Vapnik, V. (1996). Support Vector Regression Machines. *Advances in Neural Information Processing Systems*, *9*.

Ester, E. F., Anderson, D. E., Serences, J. T., & Awh, E. (2013). A Neural Measure of Precision in Visual Working Memory. *Journal of Cognitive Neuroscience*, *25*(5), 754–761. https://doi.org/10.1162/jocn_a_00357

Ester, E. F., Serences, J. T., & Awh, E. (2009). Spatially Global Representations in Human Primary Visual Cortex during Working Memory Maintenance. *Journal of Neuroscience*, *29*(48), 15258–15265. https://doi.org/10.1523/JNEUROSCI.4388-09.2009

Gazzaley, A., & Nobre, A. C. (2012). Top-down modulation: Bridging selective attention and working memory. *Trends in Cognitive Sciences*, *16*(2), 129–135. https://doi.org/10.1016/j.tics.2011.11.014

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology*, *48*(12), 1711–1725. https://doi.org/10.1111/j.1469-8986.2011.01273.x

Hallenbeck, G. E., Sprague, T. C., Rahmati, M., Sreenivasan, K. K., & Curtis, C. E. (2021). Working memory representations in visual cortex mediate distraction effects. *Nature Communications*, *12*(1), 4714. https://doi.org/10.1038/s41467-021-24973-1

Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*(7238), 632–635. https://doi.org/10.1038/nature07832

Hebart, M. N., Görgen, K., & Haynes, J.-D. (2015). The Decoding Toolbox (TDT): A versatile software package for multivariate analyses of functional imaging data. *Frontiers in Neuroinformatics*, *8*, 88. https://doi.org/10.3389/fninf.2014.00088

Iamshchinina, P., Christophel, T. B., Gayet, S., & Rademaker, R. L. (2021). Essential considerations for exploring visual working memory storage in the human brain. *Visual Cognition*, *29*(7), 425–436. https://doi.org/10.1080/13506285.2021.1915902

Jacobs, C., Schwarzkopf, D. S., & Silvanto, J. (2018). Visual working memory performance in aphantasia. *Cortex*, *105*, 61–73. https://doi.org/10.1016/j.cortex.2017.10.014

Jeffreys, H. (1998). *The theory of probability*. OUP Oxford.

Keogh, R., Bergmann, J., & Pearson, J. (2020). Cortical excitability controls the strength of mental imagery. *ELife*, *9*, e50232. https://doi.org/10.7554/eLife.50232

Keogh, R., & Pearson, J. (2014). The sensory strength of voluntary visual imagery predicts visual working memory capacity. *Journal of Vision*, *14*(12), 7–7. https://doi.org/10.1167/14.12.7

Keogh, R., Wicken, M., & Pearson, J. (2021). Visual working memory in aphantasia: Retained accuracy and capacity with a different strategy. *Cortex*, *143*, 237–253. https://doi.org/10.1016/j.cortex.2021.07.012

Keysers, C., Gazzola, V., & Wagenmakers, E.-J. (2020). Using Bayes factor hypothesis testing in neuroscience to establish evidence of absence. *Nature Neuroscience*, *23*(7), 788–799. https://doi.org/10.1038/s41593-020-0660-4

Klein, I., Dubois, J., Mangin, J.-F., Kherif, F., Flandin, G., Poline, J.-B., Denis, M., Kosslyn, S. M., & Le Bihan, D. (2004). Retinotopic organization of visual mental images as revealed

by functional magnetic resonance imaging. *Cognitive Brain Research*, *22*(1), 26–31. https://doi.org/10.1016/j.cogbrainres.2004.07.006

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception*, *36*.

Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2001). Neural foundations of imagery. *Nature Reviews Neuroscience*, *2*(9), 635–642. https://doi.org/10.1038/35090055

Kosslyn, S. M., & Thompson, W. L. (2003). When is early visual cortex activated during visual mental imagery? *Psychological Bulletin*, *129*(5), 723–746. https://doi.org/10.1037/0033-2909.129.5.723

Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, *23*(11), 571–579. https://doi.org/10.1016/S0166-2236(00)01657-X

Lee, S.-H., & Baker, C. I. (2016). Multi-Voxel Decoding and the Topography of Maintained Information During Visual Working Memory. *Frontiers in Systems Neuroscience*, *10*. https://doi.org/10.3389/fnsys.2016.00002

Lee, S.-H., Kravitz, D. J., & Baker, C. I. (2012). Disentangling visual imagery and perception of real-world objects. *NeuroImage*, *59*(4), 4064–4073. https://doi.org/10.1016/j.neuroimage.2011.10.055

Lee, S.-H., Kravitz, D. J., & Baker, C. I. (2013). Goal-dependent dissociation of visual and prefrontal cortices during working memory. *Nature Neuroscience*, *16*(8), 997–999. https://doi.org/10.1038/nn.3452

Logie, R. H., Pernet, C. R., Buonocore, A., & Sala, S. D. (2011). Low and high imagers activate networks differentially in mental rotation. *Neuropsychologia*, *49*(11), 3071–3077. https://doi.org/10.1016/j.neuropsychologia.2011.07.011

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, J., Ly, A., Gronau, Q. F., Smíra, M., Epskamp, S., Matzke, D., Wild, A., Knight, P., Rouder, J. N., Morey, R. D., & Wagenmakers, E.-J. (2019). JASP : Graphical Statistical Software for Common Statistical Designs. *Journal of Statistical Software*, *88*(2). https://doi.org/10.18637/jss.v088.i02

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Marks, D. F. (1973). Visual Imagery Differences In The Recall Of Pictures. *British Journal of Psychology*, *64*(1), 17–24. https://doi.org/10.1111/j.2044-8295.1973.tb01322.x

Mechelli, A. (2004). Where Bottom-up Meets Top-down: Neuronal Interactions during Perception and Imagery. *Cerebral Cortex*, *14*(11), 1256–1265. https://doi.org/10.1093/cercor/bhh087

Merriam, E. P., Gardner, J. L., Movshon, J. A., & Heeger, D. J. (2013). Modulation of Visual Responses by Gaze Direction in Human Visual Cortex. *Journal of Neuroscience*, *33*(24), 9879–9889. https://doi.org/10.1523/JNEUROSCI.0500-12.2013

Mongillo, G., Barak, O., & Tsodyks, M. (2008). Synaptic Theory of Working Memory. *Science*, *319*(5869), 1543–1546. https://doi.org/10.1126/science.1150769

Morey, R. D., & Rouder, J. N. (2011). Bayes factor approaches for testing interval null hypotheses. *Psychological Methods*, *16*(4), 406–419. https://doi.org/10.1037/a0024377

Moro, V., Berlucchi, G., Lerch, J., Tomaiuolo, F., & Aglioti, S. M. (2008). Selective deficit of mental visual imagery with intact primary visual cortex and visual perception. *Cortex*, *44*(2), 109–118. https://doi.org/10.1016/j.cortex.2006.06.004

Naselaris, T., Olman, C. A., Stansbury, D. E., Ugurbil, K., & Gallant, J. L. (2015). A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes. *NeuroImage*, *105*, 215–228. https://doi.org/10.1016/j.neuroimage.2014.10.018

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, *2011*, 1–9. https://doi.org/10.1155/2011/156869

Pearson, J. (2019). The human imagination: The cognitive neuroscience of visual mental imagery. *Nature Reviews Neuroscience*, *20*(10), 624–634. https://doi.org/10.1038/s41583-019-0202-9

Pearson, J. (2020). Reply to: Assessing the causal role of early visual areas in visual mental imagery. *Nature Reviews Neuroscience*, *21*, 2. https://doi.org/10.1038/s41583-020-0349-4

Pearson, J., Clifford, C. W. G., & Tong, F. (2008). The Functional Impact of Mental Imagery on Conscious Perception. *Current Biology*, *18*(13), 982–986. https://doi.org/10.1016/j.cub.2008.05.048

Pearson, J., & Keogh, R. (2019). Redefining Visual Working Memory: A Cognitive-Strategy, Brain-Region Approach. *Current Directions in Psychological Science*, *28*(3), 266–273. https://doi.org/10.1177/0963721419835210

Pearson, J., & Kosslyn, S. M. (2015). The heterogeneity of mental representation: Ending the imagery debate. *Proceedings of the National Academy of Sciences*, *112*(33), 10089–10092. https://doi.org/10.1073/pnas.1504933112

Pearson, J., Rademaker, R. L., & Tong, F. (2011). Evaluating the Mind's Eye: The Metacognition of Visual Imagery. *Psychological Science*, *22*(12), 1535–1542. https://doi.org/10.1177/0956797611417134

Pilly, P. K., & Seitz, A. R. (2009). What a difference a parameter makes: A psychophysical comparison of random dot motion algorithms. *Vision Research*, *49*(13), 1599–1612. https://doi.org/10.1016/j.visres.2009.03.019

Purdon, P. L., & Weisskoff, R. M. (1998). Effect of temporal autocorrelation due to physiological noise and stimulus paradigm on voxel-level false-positive rates in fMRI. *Human Brain Mapping*, *6*(4), 239–249. https://doi.org/10.1002/(SICI)1097-0193(1998)6:4<239::AID-HBM4>3.0.CO;2-4

Rademaker, R. L., & Pearson, J. (2012). Training Visual Imagery: Improvements of Metacognition, but not Imagery Strength. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00224

Ragni, F., Tucciarelli, R., Andersson, P., & Lingnau, A. (2020). Decoding stimulus identity in occipital, parietal and inferotemporal cortices during visual mental imagery. *Cortex*, *127*, 371–387. https://doi.org/10.1016/j.cortex.2020.02.020

Reddy, L., Tsuchiya, N., & Serre, T. (2010). Reading the mind's eye: Decoding category information during mental imagery. *NeuroImage*, *50*(2), 818–825. https://doi.org/10.1016/j.neuroimage.2009.11.084

Rose, N. S., LaRocque, J. J., Riggall, A. C., Gosseries, O., Starrett, M. J., Meyering, E. E., & Postle, B. R. (2016). Reactivation of latent working memories with transcranial magnetic stimulation. *Science*, *354*(6316), 1136–1139. https://doi.org/10.1126/science.aah7011

Senden, M., Emmerling, T. C., van Hoof, R., Frost, M. A., & Goebel, R. (2019). Reconstructing imagined letters from early visual cortex reveals tight topographic correspondence between visual mental imagery and perception. *Brain Structure and Function*, *224*(3), 1167–1183. https://doi.org/10.1007/s00429-019-01828-6

Serences, J., Saproo, S., Scolari, M., Ho, T., & Muftuler, L. (2009). Estimating the influence of attention on population codes in human visual cortex using voxel-based tuning functions. *NeuroImage*, *44*(1), 223–231. https://doi.org/10.1016/j.neuroimage.2008.07.043

Serences, J. T. (2016). Neural mechanisms of information storage in visual short-term memory. *Vision Research*, *128*, 53–67. https://doi.org/10.1016/j.visres.2016.09.010

Spagna, A., Hajhajate, D., Liu, J., & Bartolomeo, P. (2021). Visual mental imagery engages the left fusiform gyrus, but not the early visual cortex: A meta-analysis of neuroimaging evidence. *Neuroscience & Biobehavioral Reviews*, *122*, 201–217. https://doi.org/10.1016/j.neubiorev.2020.12.029

Sreenivasan, K. K., Curtis, C. E., & D'Esposito, M. (2014). Revisiting the role of persistent neural activity during working memory. *Trends in Cognitive Sciences*, *18*(2), 82–89. https://doi.org/10.1016/j.tics.2013.12.001

Tanabe, J., Miller, D., Tregellas, J., Freedman, R., & Meyer, F. G. (2002). Comparison of Detrending Methods for Optimal fMRI Preprocessing. *NeuroImage*, *15*(4), 902–907. https://doi.org/10.1006/nimg.2002.1053

Teng, C., & Kravitz, D. J. (2019). Visual working memory directly alters perception. *Nature Human Behaviour*, *3*(8), 827–836. https://doi.org/10.1038/s41562-019-0640-4

Thorudottir, S., Sigurdardottir, H. M., Rice, G. E., Kerry, S. J., Robotham, R. J., Leff, A. P., & Starrfelt, R. (2020). The Architect Who Lost the Ability to Imagine: The Cerebral Basis of Visual Imagery. *Brain Sciences*, *10*(2), 59. https://doi.org/10.3390/brainsci10020059

Tong, F. (2013). Imagery and visual working memory: One and the same? *Trends in Cognitive Sciences*, *17*(10), 489–490. https://doi.org/10.1016/j.tics.2013.08.005

Töpfer, F. M., Barbieri, R., Sexton, C. M., Wang, X., Soch, J., Bogler, C., & Haynes, J.-D. (2022). Psychophysics and computational modeling of feature-continuous motion perception. *Journal of Vision*, *22*(11), 16. https://doi.org/10.1167/jov.22.11.16

Urai, A. E., Braun, A., & Donner, T. H. (2017). Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. *Nature Communications*, *8*(1), 14637. https://doi.org/10.1038/ncomms14637

Wang, L., Mruczek, R. E. B., Arcaro, M. J., & Kastner, S. (2015). Probabilistic Maps of Visual Topography in Human Cortex. *Cerebral Cortex*, *25*(10), 3911–3931. https://doi.org/10.1093/cercor/bhu277

Zeman, A., Dewar, M., & Della Sala, S. (2015). Lives without imagery – Congenital aphantasia. *Cortex*, *73*, 378–380. https://doi.org/10.1016/j.cortex.2015.05.019

Zeman, A. Z. J., Della Sala, S., Torrens, L. A., Gountouna, V.-E., McGonigle, D. J., & Logie, R. H. (2010). Loss of imagery phenomenology with intact visuo-spatial task performance: A case of 'blind imagination.' *Neuropsychologia*, *48*(1), 145–155. https://doi.org/10.1016/j.neuropsychologia.2009.08.024

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233–235. https://doi.org/10.1038/nature06860

Extended Figures

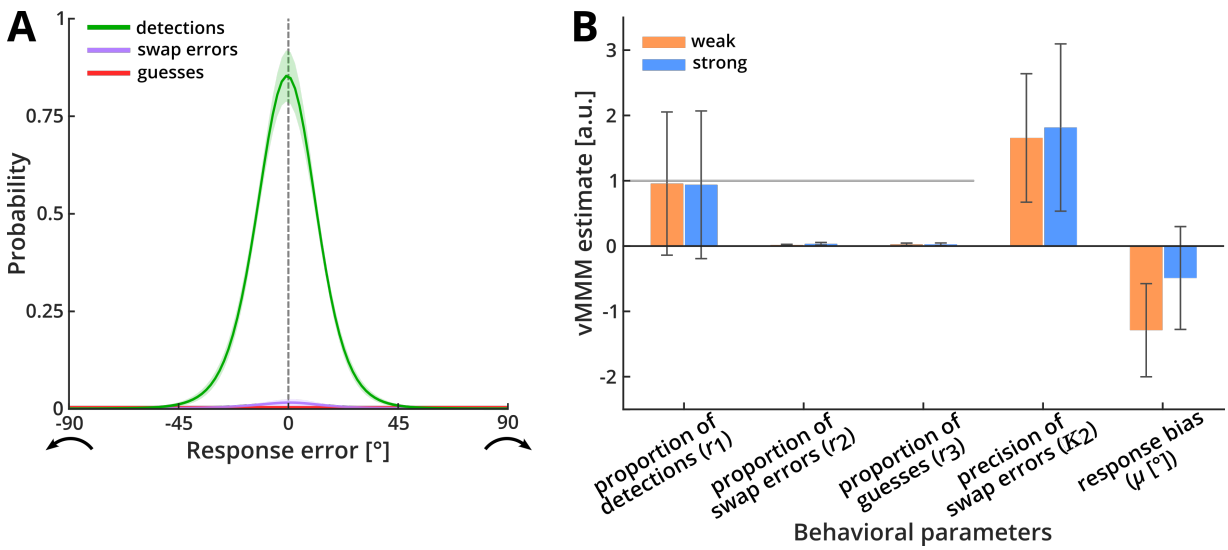

Fig. S1: Von Mises mixture model (vMMM) fit of behavioral responses. (A) The distribution of behavioral responses was modeled as a combination of the three model components: detections (responses to target orientations, assumed to follow a von Mises distribution with mean 0° plus bias μ and precision κ; green), swap errors (responses to distractor orientations, following the same assumptions as detections; purple) and guesses (assumed to follow a continuous uniform distribution between -90° and +90°; red). These components were weighted by individual event probabilities (mixture coefficients) r_1, r_2 and r_3, respectively. Participants correctly responded to the target direction in 94.7 % of trials (r_1 = 0.947 ± 0.063), and only infrequently made swap errors (r_2 = 0.026 ± 0.034) or guesses (r_3 = 0.027 ± 0.041). Responses to targets were precise (κ_1 = 5.673 ± 2.377), while responses to the distractor, where present, were imprecise (κ_2 = 1.735 ± 2.41). There was a small but significant bias to respond anti-clockwise of the target (μ = -0.889 ± 1.635°; $t_{(39)}$ = -3.437, p = 0.0014, two-tailed; see also Fig. 1C). **(B)** Estimated vMMM parameters for strong and weak imagers separately. There was no significant difference between the two groups for any of the estimated parameters (r_1: $t_{(38)}$ = -0.925, p = 0.361; r_2: $t_{(38)}$ = 1.585, p = 0.121; r_3: $t_{(38)}$ = 0.108, p = 0.914; κ_2: $t_{(38)}$ = -0.207, p = 0.837; μ: $t_{(38)}$ = 1.574, p = 0.124, all two-tailed; see Fig. 1D for κ_1).

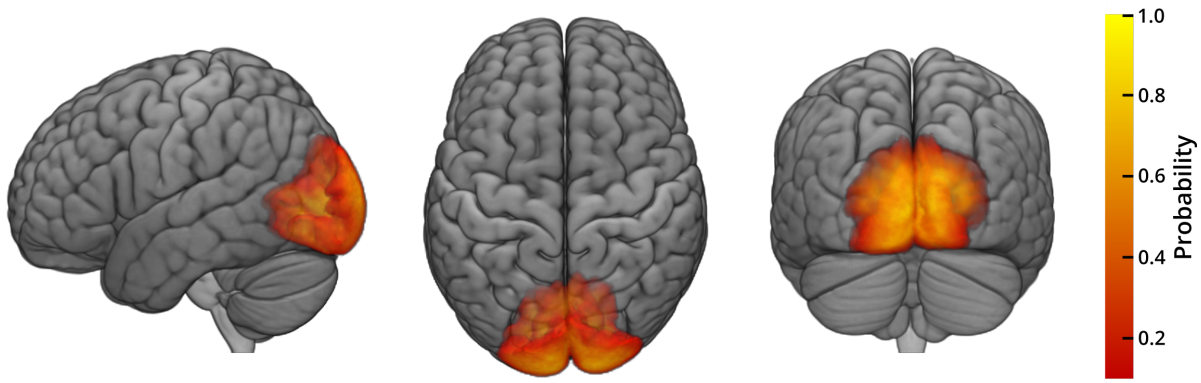

Fig. S2: Probabilistic map of V1-3 used for voxel selection. We combined probabilistic anatomical maps of V1, V2 and V3 (Wang et al., 2015) to create a bilateral ROI of early visual cortex in standard space. The ROI was used for voxel selection in the reconstruction analysis: First, the ROI was transformed into the native space of each participant and thresholded at 0.1. Then, we used the intersection of the transformed maps and native stimulation-vs.-baseline contrasts to select the most activated voxels that fell into this ROI for each participant. The precise number of voxels used for reconstruction was subject to a nested cross-validation across participants.

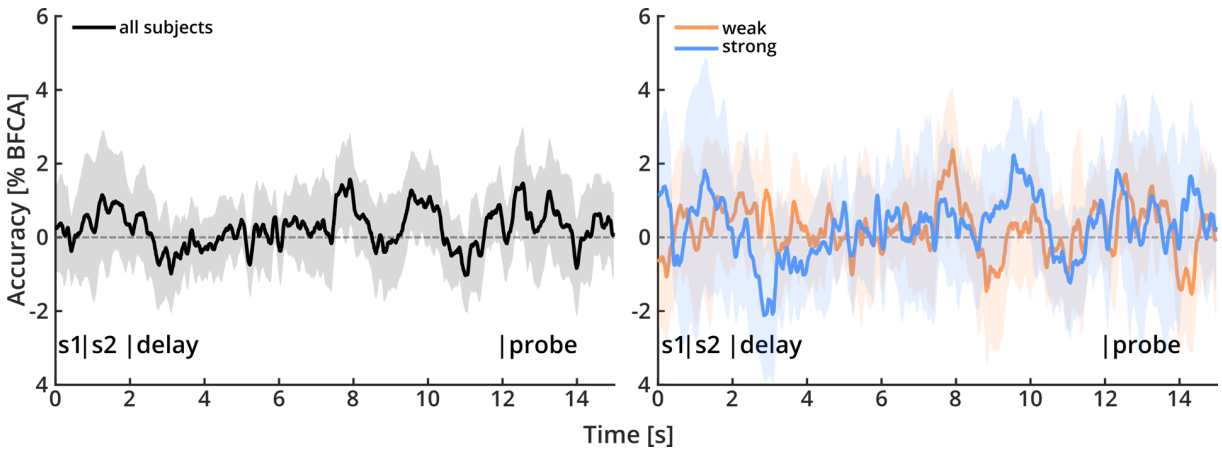

Fig. S3: Target reconstruction from eye-tracking data. Reconstruction of target orientation from gaze position across the trial, for all subjects (left panel) and separated by groups (right panel). There were no temporal clusters with significantly above-chance BFCA, suggesting that participants did not systematically use gaze position to maintain target orientation across the delay period. Shaded areas indicate 95 % confidence intervals.

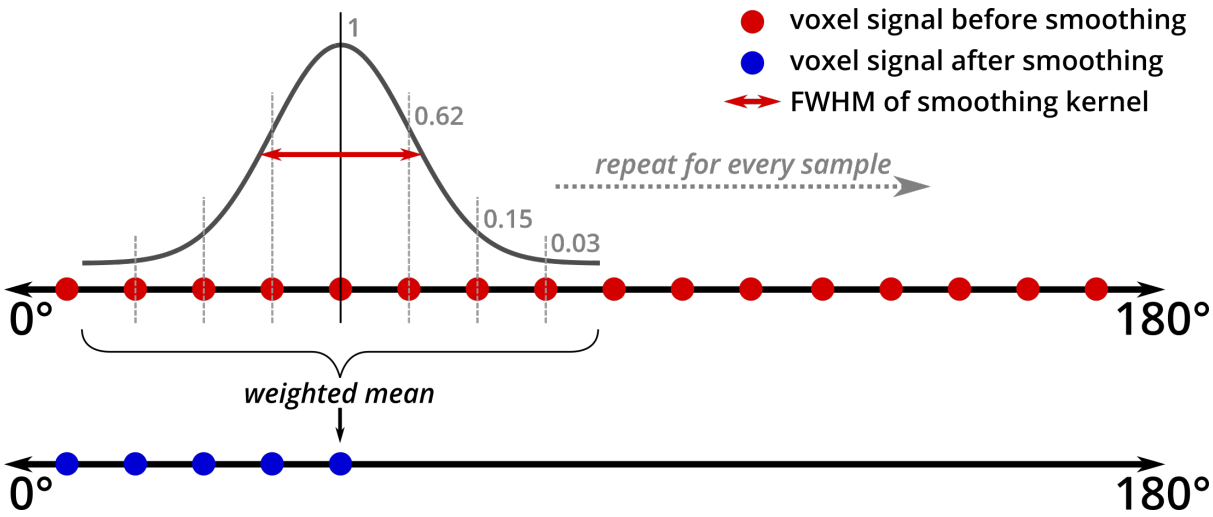

Fig. S4: Schematic representation of feature-space smoothing. We used a Gaussian smoothing kernel to compute a weighted average from the voxel signal of samples lying closely together in feature space. Samples close to a given orientation in feature-space therefore contribute more to the resulting average than those further away. The full width at half maximum (FWHM) of the Gaussian kernel controls the smoothing range, i.e., the number (or distance) of samples that are included in the weighted average. We used FWHM values between 0° (no smoothing) and 90° in steps of 10° and determined the optimal kernel width for each participant via nested cross-validation across subjects. Note that this was done (a) at the level of the input data to the analysis, not the results, (b) for training and test data separately, and (c) was confirmed not to produce artifacts or spurious results by extensive simulations (see Fig. S5 and Extended Methods).

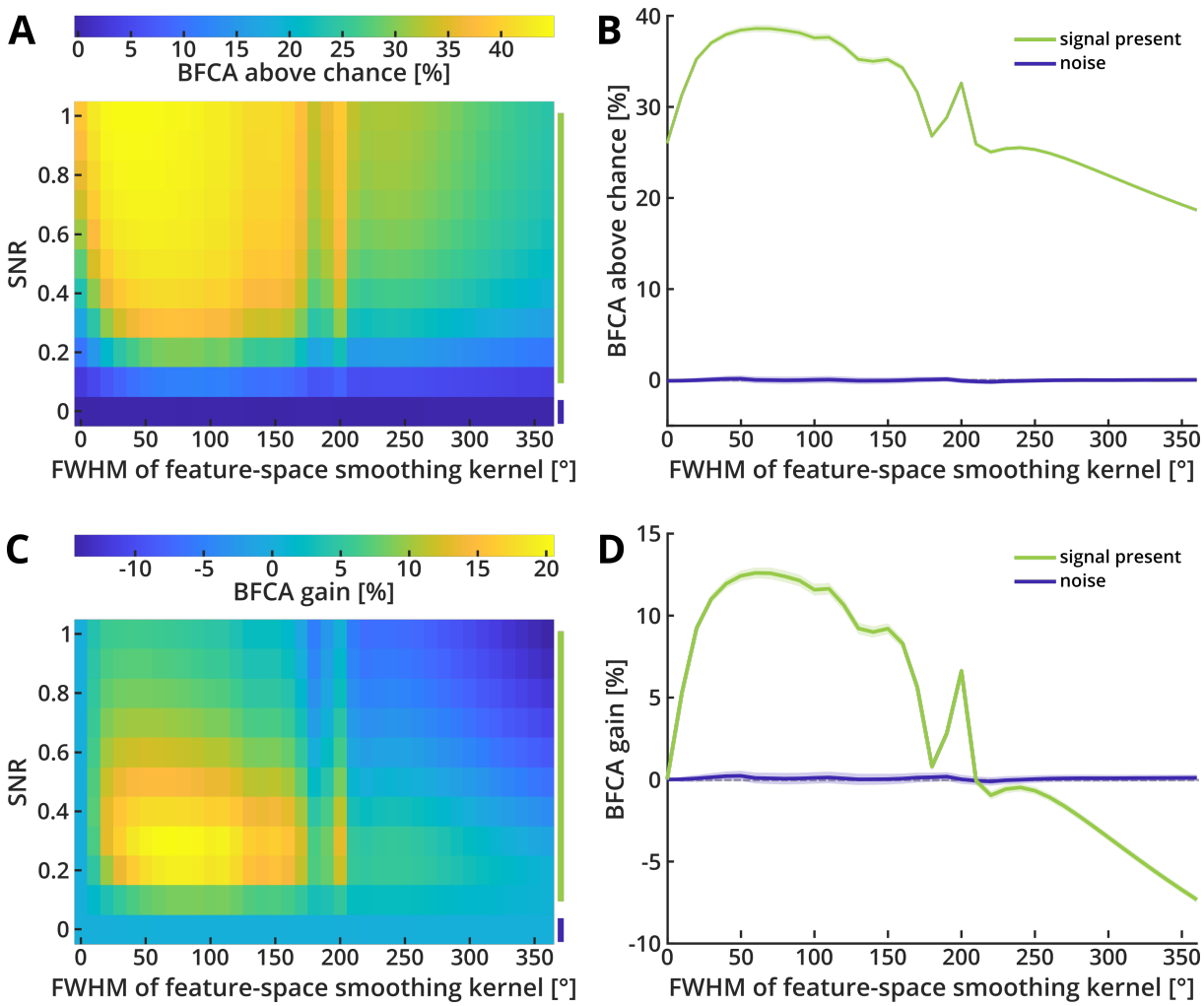

Fig. S5: Feature-space smoothing simulation results. We simulated data with varying levels of SNR and used feature-space smoothing with different kernel widths (measured as FWHM in degrees) before reconstruction of the underlying signal. **(A)** BFCA for all parameter combinations, averaged across 1000 repetitions. **(B)** BFCA across smoothing levels, for pure noise (dark blue) and averaged across all signal conditions (SNR = 0.1 and above, green). **(C)** BFCA gain compared to no smoothing, averaged across all 1000 repetitions. The first column corresponds to baseline, i.e., zero smoothing. BFCA gain was especially pronounced for FWHM values between 30° and 170° in cases of low SNR. **(D)** BFCA gain across feature-space smoothing levels, for pure noise (dark blue) and averaged across all signal conditions (SNR = 0.1 and above, green).

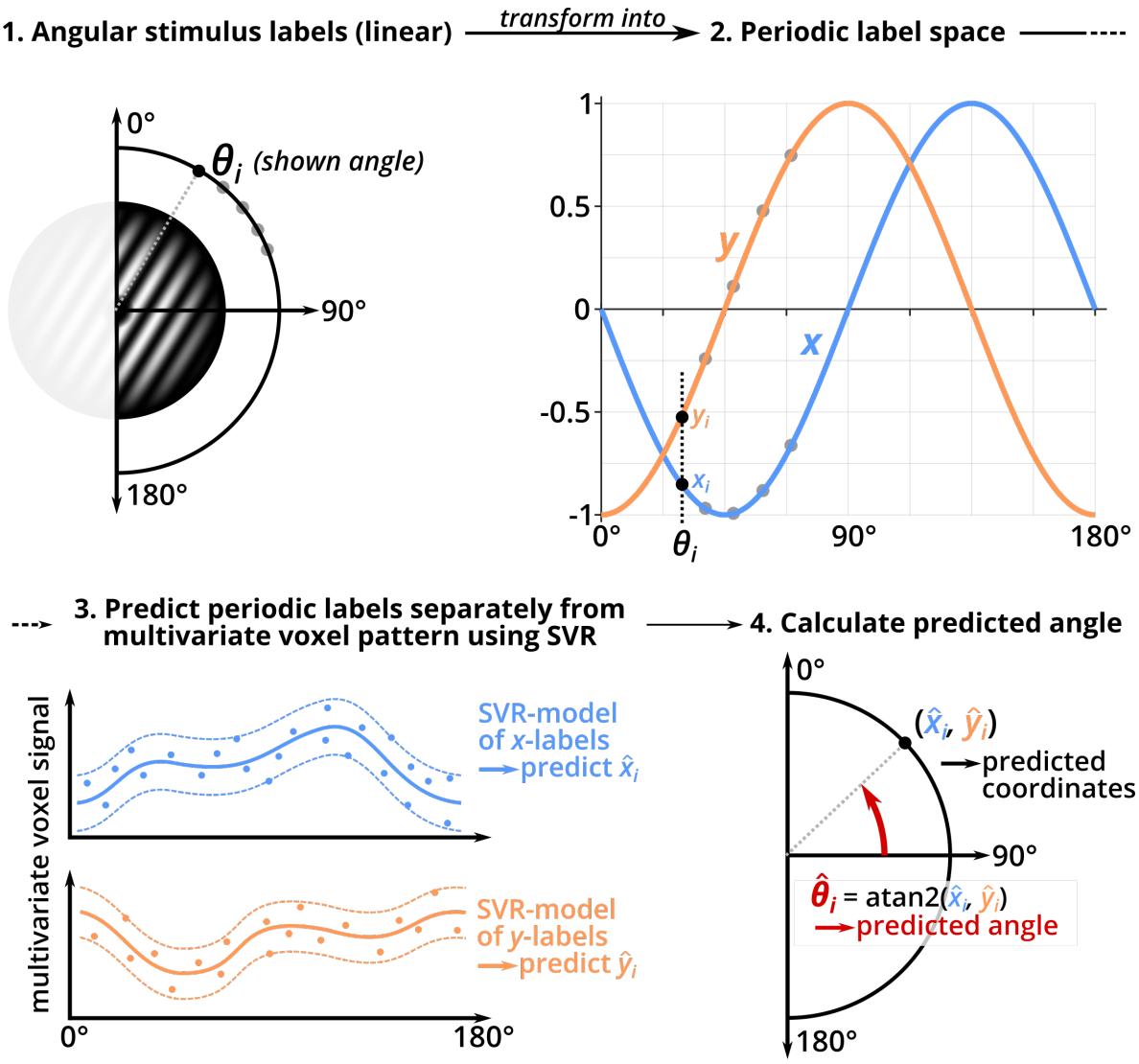

Fig. S6: Schematic representation of periodic support vector regression (pSVR). The aim of our reconstruction analysis was to predict an angular label between 0° and 180° from the multivariate voxel signal in response to a stimulus grating with the respective orientation. However, the linear scale of orientation labels (from 0° to 180°) does not reflect the periodic nature of the stimulus (i.e., 0° and 180° are identical). To account for this, we projected the angular labels into a periodic space by fitting two sinusoids into the range [0, 180). Both functions had an amplitude of 1 and a period of 180°, so that one period spanned the entire label space. One function was shifted by 45°, so that the combination of both periodic functions coded for the linear label scale. This is equivalent to the way sine and cosine functions between 0 and 360° code for the angles on a unit circle. We trained and tested a multivariate SVR model for both periodic label sets (*x, y*) separately. From the combination of the predicted periodic labels, we then reconstructed a predicted angular label using the four-quadrant inverse tangent. The predicted orientation was then compared to the true orientation to derive BFCA, our measure of reconstruction accuracy.

Table S1: Results for group splits based on post-scan questionnaires.

Questionnaire type	t_{38}	*p*	BF_{01} (high ≠ low)	BF_{01} (high > low)
VVIQ post-scan	-0.956	0.345	2.255	5.629
OSIQ (visual)	0.129	0.898	3.217	2.949
OSIQ (spatial)	1.365	0.18	1.555	0.879
Strategy (visual)	0.495	0.624	2.937	2.197
Strategy (verbal)	1.071	0.291	2.056	1.243

To verify the results from our main analysis which relied on groups derived from the recruitment VVIQ scores (Fig. 2B/C), we re-assigned the groups based on various post-scan questionnaire scores (see Methods). For each questionnaire, we sorted the respective scores from low to high and assigned the participants corresponding to the bottom half to the weak imagery group, and those corresponding to the top half to the strong imagery group. We observed no significant group differences for any of the post-scan splits, both with conventional and Bayesian *t*-tests, thus confirming the results we obtained using the recruitment VVIQ scores. Therefore, we judge the absence of a group effect in our main analysis as a general pattern in our data, rather than an artifact of the recruitment scores specifically. Please note that we report the Bayes factor in favor of the null hypothesis: the higher the value, the less evidence for the alternative hypothesis.

Extended Methods

Feature-space smoothing simulation

To demonstrate how feature-space smoothing can increase signal-to-noise ratio (SNR) and increase accuracy in a continuous reconstruction setting, we simulated fMRI data with varying amounts of SNR and used different levels of feature-space smoothing before reconstruction. Following the specifics of our experiment, we simulated data comprising 8 runs with 40 trials each, for 250 voxels. The measured response of voxel *i* in trial *j* was generated as

$$d_{ij} = y_{ij} \cdot s + \varepsilon_{ij}$$

where y_{ij} is the actual response of voxel *i* in response to the orientation shown in trial *j*, *s* is a scaling factor controlling the ratio of signal and noise, and ε_{ij} is sampled from a standard normal distribution.

To simulate the voxel responses, we assumed a population of idealized voxels, where each voxel would exhibit a distinct periodic tuning profile in response to angular orientation. The tuning profile z_i for each voxel *i* was sampled from a multivariate normal distribution

$$z_i \sim N(0, K_i)$$

where K_i specifies the voxels' periodic covariance kernel. This kernel K_i is given by

$$K_i(x_m, x_n, \sigma_i) = exp\left[-2\left(\frac{sin\left(\frac{1}{2}(x_m - x_n)\right)}{\sigma_i}\right)^2\right]$$

where *x* is a $p \times 1$ vector specifying a grid of possible orientations, such that $x_m, x_n \in [0,2\pi)$, p is controlling the number of unique, equally spaced values from the feature space; and σ_i is the voxel's unique tuning function smoothness parameter. For this simulation, the smoothness of each voxel was sampled from a gamma distribution: $\sigma_i \sim \Gamma(2,2)$.

Thus, voxel- and trial-wise responses could be sampled as

$$y_{ij} = z_i(x_j)$$

where x_j is the orientation presented during the *j*-th trial and orientation labels were drawn from a uniform distribution: $x_j \sim U(0,2\pi)$.

For the SNR-controlling factor *s*, we chose 10 values between 0.1 and 1, equally spaced by 0.1, as well as 0 (i.e., pure noise). Before reconstruction, we used feature-space smoothing on the data, for FWHM values between 0° (i.e., no smoothing) and 360°, equally spaced by 10°. This resulted in 11 SNR levels and 37 smoothing levels. After pSVR reconstruction, we calculated BFCA as our measure of accuracy. The simulation was repeated 1000 times for each parameter combination.

First, we wanted to verify that feature-space smoothing increases reconstruction accuracy specifically by improving the SNR in the data, and not by introducing artifacts that might be exploited by the reconstruction model. Therefore, we tested for the noise condition ($s = 0$) whether (i) BFCA would differ from chance-level for any smoothing level (using one-sample two-tailed *t*-tests) and (ii) BFCA from any smoothing level would differ from the no-smoothing condition (using two sample two-tailed *t*-tests). We show that (i) BFCA remained at chance-level across all levels of smoothing (all $p > 0.25$, Fig. S5B) and (ii) BFCA for any smoothing condition did not differ from the no-smoothing condition (all $p > 0.15$, Fig. S5D).
Moreover, In the signal conditions ($s \neq 0$), feature-space smoothing was able to reliably increase BFCA compared to no smoothing. The effect was strongest for smoothing kernel widths between 30° and 170°, where we observed increases in accuracy of up to 20 % (Fig. S5C). Generally, the effect of feature space smoothing was stronger for data with low SNR (Fig. S5C, orange-yellow area). In cases of extremely high kernel-width and comparatively high SNR (i.e., SNR > 0.6 and FWHM > 220°), feature-space smoothing had a detrimental effect, meaning that BFCA was decreased compared to no smoothing (Fig. S5C, dark blue area). Please note, however, that kernel-widths this high do not make any sense for real-world applications and were only included for the purpose of demonstration. Based on these results, we conclude that feature-space smoothing is a powerful preprocessing technique to increase SNR in a feature-continuous reconstruction setting. As the optimal kernel-width for smoothing depends on the specific data and SNR, we recommend using nested cross-validation to determine the optimal FWHM value, similar to the approach described in the main text.

Reconstruction of feature-continuous orientation from eye-tracking data

Since we wanted our participants to employ *cognitive, not motor-based* strategies to remember the target orientations, they were explicitly instructed to maintain fixation at all times during the experiment. However, it is still possible that participants use eye movements towards a certain direction as a *non-cognitive* strategy to remember target orientations. Furthermore, eye movements have been shown to modulate visual responses in the brain (Merriam et al., 2013). To account for these potentially confounding factors, we investigated whether the gaze position across the trial held information about the target orientation, by subjecting the recorded x and y ordinates of 26 participants to the same reconstruction analysis as the fMRI data.

Preprocessing of eye-tracking data was performed in MATLAB using functions from the *Fieldtrip* toolbox (Oostenveld et al., 2011), code adapted from prior work (Urai et al., 2017) and in-house code. Blinks were linearly interpolated and bandpass filtered between 5 Hz (high-pass) and 100 Hz (low-pass). For each trial, we extracted 15 s worth of data following the onset of the first grating. The data from each run was detrended using the same cubic spline interpolation as described above (see *Preprocessing of fMRI data*). We then downsampled the data by a factor of 10, resulting in 1500 time points per trial.

After preprocessing, we entered the data into the same pSVR reconstruction analysis as the fMRI data, using the x and y ordinates of the gaze position as features instead of voxel signals, and evaluated the reconstruction by calculating the BFCA. As with the fMRI data, we tested for clusters of above-chance time points using the cluster-based *t*-mass permutation approach described above.