NeuroImage xxx (2017) 1–10

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Decoupling of BOLD amplitude and pattern classification of orientation-selective activity in human visual cortex

Anke Marit Albers [*], Thomas Meindertsma [1,2], Ivan Toni, Floris P. de Lange

Radboud University, Donders Institute for Brain, Cognition and Behaviour, P.O. Box 9101, NL-6500 HB Nijmegen, The Netherlands

ARTICLE INFO

Keywords:
BOLD activity
MVPA decoding
Early visual cortex
Orientation specificity

ABSTRACT

Multivariate pattern analysis (MVPA) of fMRI data has allowed the investigation of neural representations of stimuli on the basis of distributed patterns of activity within a brain region, independently from overall brain activity. For instance, several studies on early visual cortex have reported reliable MVPA decoding of the identity of a stimulus representation that was kept in working memory or internally generated, despite the fact that the overall BOLD response was low or even at baseline levels. Here we ask how it is possible that reliable stimulus information can be decoded from early visual cortex even when the overall BOLD signal remains low. We reanalyzed a data set in which human participants (N = 24) imagined or kept in working memory an oriented visual grating. We divided voxels from V1, V2, and V3 into groups based on orientation preference, and compared the time course of mean BOLD responses to preferred and non-preferred orientations with the time course of the multivariate decoding performance.

Decoding accuracy related to a numerically small, but reliable univariate difference in the mean BOLD response to preferred and non-preferred stimuli. The time course of the *difference* in BOLD responses to preferred and non-preferred orientations was highly similar to the time course of the multivariate pattern classification accuracy. The reliability of the classification strongly correlated with the magnitude of differences in BOLD signal between preferred and non-preferred stimuli. These activity differences were small compared to the large overall BOLD modulations. This suggests that a substantial part of the task-related BOLD response to visual stimulation might not be stimulus-specific. Rather, stimulus-evoked BOLD signals in early visual cortex during a task context may be an amalgam of small stimulus-specific responses and large task-related but non-stimulus-specific responses. The latter are not evident during the maintenance or internal generation of stimulus representations, but provide an explanation of how reliable stimulus information can be decoded from early visual cortex even though its overall BOLD signal remains low.

1. Introduction

Over the past decade, multivariate pattern analysis (MVPA) has gained popularity as a sensitive method to investigate representations in the brain. By making use of the distributed patterns of activity within a region (e.g.Haynes and Rees, 2006; Kamitani and Tong, 2005), or across the entire brain (e.g. Cukur et al., 2013; Kriegeskorte et al., 2006; Mitchell et al., 2008), MVPA has allowed researchers to show that visual regions contain perception-like representations of internally generated content during (working) memory, mental imagery and even dreaming (Albers et al., 2013; Bosch et al., 2014; Harrison and Tong, 2009; Horikawa et al., 2013; Kamitani and Tong, 2006; Naselaris et al., 2015; Serences et al., 2009). In many of these studies, there was reliable decoding of stimulus identity, even though the overall BOLD response in these regions was close to, or even at, baseline levels (e.g. Albers et al., 2013; Bosch et al., 2014; Harrison and Tong, 2009; Serences et al., 2009; Sneve et al., 2012).

The classical mass-univariate approach (usually analyzed using a General Linear Model (GLM)) is sensitive to a univariate code: an effect that has the same direction in all voxels considered (Woolgar et al., 2014). MVPA, on the other hand, gains its sensitivity from taking into account changes in activity in multiple voxels, regardless of the direction of change – information that is usually lost when averaging. MVPA can be sensitive to a multivariate (neural) code, i.e. a pattern of interactions

* Corresponding author. Current Address: Abteilung Allgemeine Psychologie, Justus-Liebig-Universität, Otto-Behaghel-Str. 10, 35394 Giessen, Germany.
E-mail addresses: ankemarit.albers@gmail.com (A.M. Albers), t.meindertsma@uva.nl (T. Meindertsma), ivan.toni@donders.ru.nl (I. Toni), floris.delange@donders.ru.nl (F.P. de Lange).
[1] Department of Psychology, University of Amsterdam, Nieuwe Achtergracht 129B, 1018 WS Amsterdam, The Netherlands.
[2] Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg- Eppendorf, Martinistrasse 52, 20246 Hamburg, Germany.

https://doi.org/10.1016/j.neuroimage.2017.09.046
Received 12 February 2017; Accepted 21 September 2017
Available online xxxx

Abbreviations

WM	working memory; the internal maintenance of information over a delay period
IM	mental imagery; the internal manipulation of information or the generation of novel representations

between the activity in several voxels, where the stimulus cannot be read out from the activity in one or several voxels but only when looking at activation and deactivation in multiple voxels combined. However, MVPA is also sensitive to univariate differences at the voxel level: when voxels show a higher response to one stimulus, a classifier can pick up on this (Norman et al., 2006).

Pattern classifiers investigate patterns of BOLD activity and linear classifiers are successful only when there are voxels showing a difference between the stimuli (Anderson and Oates, 2010). Given that MVPA relies on BOLD activity, an increased BOLD signal may be expected to go hand in hand with increases in decoding accuracy. This is indeed generally found for decoding of bottom-up visual input (Smith et al., 2011; Tong et al., 2012). Yet, studies that investigated representations of internally generated content observed no such relationship with overall BOLD signal (Harrison and Tong, 2009; Riggall and Postle, 2012). Thus, it has remained a puzzle how there can be reliable patterns of BOLD activity while overall BOLD activity is not above baseline.

There are several plausible explanations for this dissociation. First of all, the decoding results might indicate information that is present in multiple voxels, whereby the covariance of the noise can help the classifier discriminate responses. The overall level of activity might sum up to zero because active and inactive voxels might cancel each other out, making the decoding pattern independent from the overall level of BOLD activity. An appropriate *multivariate pattern analysis* might be able to discriminate between such multivariate *codes*, whereas a traditional *mass-univariate analysis* would be at loss when overall activity is equal between conditions.

Similarly, increased decoding accuracy has been linked to a decreased, but more specific, overall response (Kok et al., 2012). In the case of working memory maintenance, stimulus-specific cerebral representations (i.e. voxels) might remain active while concurrently other stimulus representations, or voxels, are inhibited, thus decreasing the overall signal even though some voxels show a strong univariate response.

Finally, as the BOLD signal is an amalgam of stimulus-specific and stimulus-unspecific signals (Donner et al., 2008; Jack et al., 2006; Kloosterman et al., 2015; Swallow et al., 2012), stimulus-specific signals might be hidden or counteracted by non-specific effects. Non-specific signals in early visual cortex have been observed for task transitions (Jack et al., 2006), temporal selection (Swallow et al., 2012) and decision-related processing (Donner et al., 2008; Kloosterman et al., 2015; Warren et al., 2015). Signal fluctuations induced by collateral task features might obscure BOLD fluctuations in stimulus-specific patterns relevant for decoding – especially when the latter effects are comparatively small. This could then result in an apparent dissociation between univariate (no activity differences) and multivariate (decodable stimulus information) results.

Here we investigate the relationship between overall and voxel-specific BOLD signals for internally generated representations. We reanalyzed a dataset (Albers et al., 2013) in which we could read out the identity of a stimulus representation that was kept in working memory or internally generated, from patterns of activity in early visual cortex. Our reanalysis demonstrates that while the overall BOLD signal modulation was much larger than the voxel-specific signal differences induced by the different stimuli, stimulus-specific BOLD differences were reliably detectable and strongly correlated with stimulus decoding ability. This suggests that internally generated and maintained stimulus representations give rise to subtle but reliable increases in BOLD signal, providing a basis for pattern classification of stimulus content independent of overall levels of BOLD activation.

2. Materials & methods

2.1. Participants and task

In the current study we reanalyzed a dataset that was previously published (Albers et al., 2013). Participants gave written informed consent in accordance with the institutional guidelines of the local ethical committee (CMO region Arnhem-Nijmegen, The Netherlands) and were paid for their participation.

In brief, 24 participants (10 males, ages 18–30), performed a mental imagery (IM) and a working memory (WM) task while in the fMRI scanner (3T Trio MRI System, $3 \times 3 \times 3$ mm voxels, TR = 2 s). During this task (Fig. 1A), participants were first shown two orientated gratings and two cues, which together instructed them about the imagery or working memory task to perform, as well as which oriented grating to use. Subsequently, they performed the task and maintained, or mentally rotated and imagined, a specific oriented grating for about 10 s. Each trial ended with the presentation of a visual probe that was rotated slightly clockwise or counterclockwise with respect to the imagined or maintained orientation. Participants decided whether the probe was rotated in clockwise or counterclockwise direction and responded with a button press followed by feedback about whether they were correct on that current trial. There were three unique orientations (15, 75 and 135°) that participants were to imagine, but due to the different rotation operations (60 and 120° rotation, clockwise and counterclockwise), and the retro-cueing procedure there were 60 unique trial types and 120 trials in total.

After the main task, participants underwent a functional localizer scan during which they were presented with the same three orientated gratings as in the main task (15, 75 and 135°). The gratings were presented for 12 s while flashed at 4 Hz, and participants pushed a response button whenever the fixation dot changed color. Finally, they underwent a scan for retinotopic mapping using a rotating wedge (Engel et al., 1997).

2.2. GLMs and selection contrasts

All analyses were performed on an individual, per subject basis. Individual data were realigned and co-registered using SPM8 (http://www.fil.ion.ucl.ac.uk/spm, Wellcome Trust Centre for Neuroimaging, London, UK). We obtained activity contrasts to use for voxel selection for individual participants using the first level analysis in SPM8. For the localizer, we used three regressors to model the activity for each orientation (15, 75 and 135°, duration 12 s) separately. For the main task we modeled the stimuli and cues (duration = 1.9 s), imagery period (duration = 10.4 s) and probe with response (duration = 2.3 s) for the WM trials and IM trials separately. We created the following contrasts:

1) *Visual stimulation contrast:* a *t*-contrast of the activity during unattended perception (functional localizer) of all oriented gratings versus baseline, regardless of orientation. This allowed us to investigate which voxels responded best to our stimuli. This was also the contrast we used for voxel selection here and in the decoding analyses (see 2.4).
2) *Task stimulus contrast*: a *t*-contrast from the main task, with the activity during the instruction cues and grating stimuli (WM and IM trials combined) against the implicit baseline. This contrast is again for voxels with the strongest response to the stimuli, but this time in the context of the main task. This contrast corresponds to the activity in the first BOLD peak in Fig. 2A and B.
3) *Task probe contrast: t*-contrast from main task, with the activity during the probe presentation (WM and IM trials combined), compared to the implicit baseline. This contrast corresponds to the activity in the second BOLD peak in the time course (Fig. 2A and B) and allowed plotting the spatial distribution of the overall activity at that point.

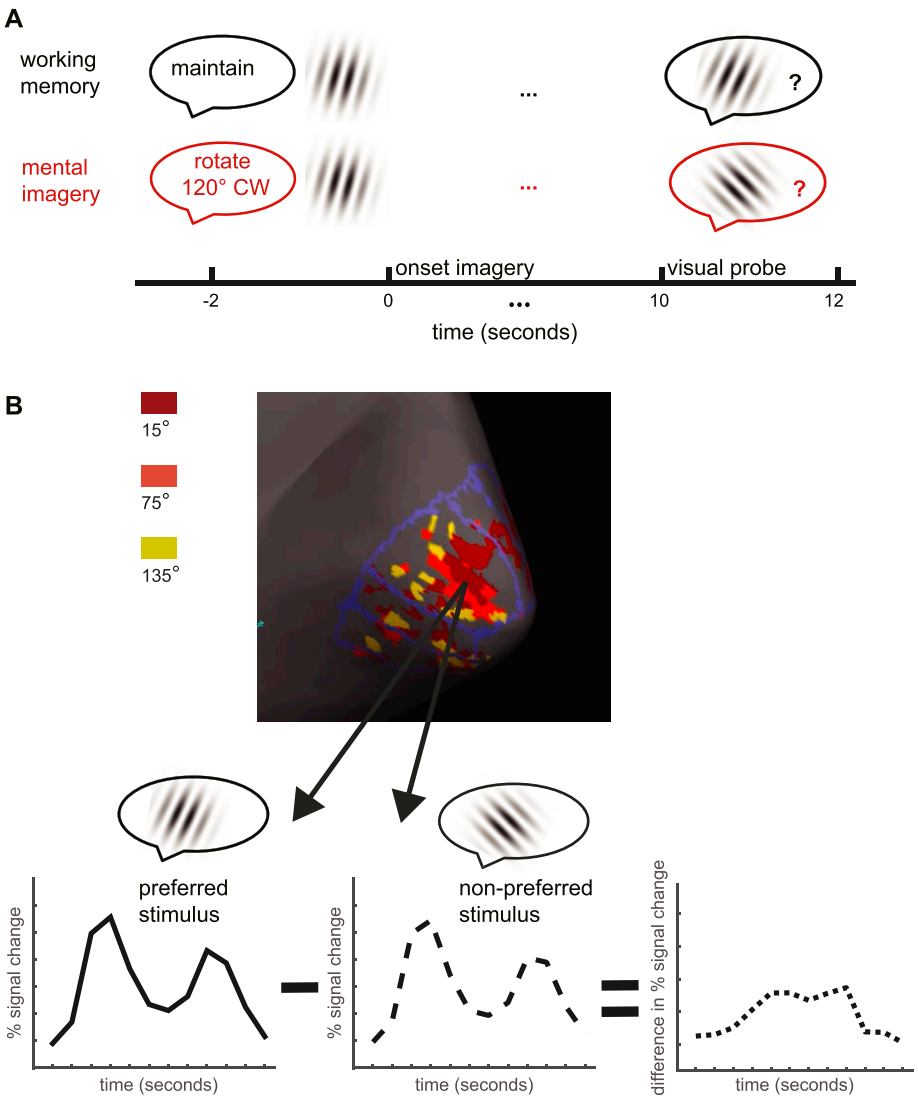

Fig. 1. Task design and analytical approach. A) Schematic of experimental design. At the start of each trial, a task cue indicated whether participants had to maintain a stimulus in working memory (WM; top row) or create a new stimulus representation by imagining rotating the oriented grating and keeping the ensuing mental image in their mind's eye (mental imagery [IM]; bottom row). During IM trials, mental rotation could be clockwise or counterclockwise and 60° or 120°. After the task cue, two gratings (out of three possible stimuli: 15°, 75°, or 115°) were presented briefly (only one shown), followed by a second stimulus cue (not shown) that indicated which stimulus grating to select and maintain (WM) or rotate and then imagine (IM). After a 10 s delay period in which participants were asked to vividly imagine the relevant orientation, a probe was presented. Participants indicated whether the probe was rotated clockwise or counterclockwise with respect to the stimulus representation they had kept in mind and received feedback on each trial. For a full figure of the task design, see (Albers et al., 2013). **B)** Analytical approach. The surface shows voxel selection: we identified the 120 most active voxels (during a separate visual stimulation localizer) in V1, V2 and V3; here plotted on the inflated surface of the right hemisphere of one representative participant. For each voxel, we assigned a preferred orientation – 15° (red), 75° (orange) or 135° (yellow); the distribution of preferences for this participant is visible in the three different colors in the plot. For each voxel, we extracted the activity time courses separately for each orientation. We averaged the activity time courses over all voxels, separately for trials at which their preferred orientation was remembered/imagined (left plot) and for trials at which the non-preferred orientations were remembered/imagined (middle plot). Subsequently, we calculated the difference between activity for preferred and non-preferred orientations for each time point, which gave us a 'difference time course' (right plot).

Regions of significant activation were determined with a threshold of $t = 4.7$, which was based on the $p < 0.05$ Family-wise error (FWE) corrected threshold as obtained from SPM8.

2.3. Retinotopy

We used Freesurfer (http://surfer/nmr/mgh/harvard/edu/) to inflate the anatomical volume of each participant and to draw the borders of V1, V2 and V3 (DeYoe et al., 1996; Engel et al., 1997; Sereno et al., 1995; Wandell et al., 2007). Regions of interest (ROIs) were created for each early visual area. Due to confluent foveal representations for V1-V3 (Wandell et al., 2007), we excluded the foveal representation from our regions of interest.

The individually drawn regions only covered the central portions of V1-V3, as the wedge used for retinotopic mapping was only slightly larger than the grating stimuli in the main experiment. To investigate BOLD modulations in regions of visual cortex not directly stimulated by the gratings, we therefore created a second set of ROIs that included the more peripheral parts of the visual regions. These ROIs were based on the V1, V2 and V3 templates from Benson and colleagues (Benson et al., 2012, 2014, https://cfn.upenn.edu/aguirre/wiki/public:retinotopy_template), that can be applied to individual brains based on anatomy. A major advantage of these templates is that they not only contain labels for each area, but also contain information about eccentricity (up to 90 degrees of visual angle) as well as retinotopic angle. We converted these templates to match the functional brain scans of each participants and used them to select voxels for the eccentricity analysis in central and peripheral early visual cortex. We defined peripheral V1 as those voxels that had an eccentricity >15 degrees of visual angle.

2.4. Voxel selection

Within each central ROI, the 120 most active voxels during perception of all three gratings (during the independent functional localizer) were selected (120 most active voxels in the *visual stimulation contrast*, see 2.2). For the combined early visual cortex ROI, all the voxels from the individual V1-V3 ROIs were selected ($3 \times 120 = 360$ voxels total). We extracted the BOLD time course for each voxel in the ROIs and high pass-filtered the data (removing signal with $f < 1/128$ Hz) and detrended to remove slow drifts during the scanner runs.

2.5. Decoding analyses

In all decoding analyses, we trained linear support vector machines (SVMs) to discriminate between the three grating orientations based on the pattern of BOLD activity over voxels during the localizer (unattended perception). We then applied the SVMs to the imagery and working memory data to classify imagined orientation from the activity patterns during mental imagery and working memory maintenance for each scan (for details, please see (Albers et al., 2013)).

2.6. Separating voxels based on their preference

To obtain time courses for preferred and non-preferred orientations, we divided the voxels according to whether they preferred the imagined orientation or one of the other orientations, and then plotted the mean BOLD amplitude difference between preferred and non-preferred voxels over time (Fig. 1B). First, we assigned a preferred orientation to every selected voxel by comparing the *t*-values from the following contrasts: **15** vs. 75 & 135, **75** vs. 15 & 135 and **135** vs. 15 & 75. By using these contrasts, we were able to obtain voxels that were most *selective* to one of the orientations. The contrast that yielded the highest *t*-value for that voxel was taken as the preferred orientation and assigned to the voxel. Every voxel could only be assigned one orientation (Fig. 1B). Per voxel group, we then separated the trials with the preferred orientation for that group (1/3 of the trials) from the trials with a non-preferred orientation and averaged, per time point, over voxels, trials and scanner runs (6 runs total), but for preferred and non-preferred orientation trials separately. At this point in the analysis we had 3 'orientation-preference'-groups with each one standardized mean BOLD time course for representation of their preferred stimulus and one for the non-preferred stimuli. We then averaged over voxel groups for preferred and non-preferred time courses independently. As a control, we tested whether there were any a-priori activity differences between the voxel groups. We averaged the activity over all selected voxels in each ROI for each orientation averaged and performed a one-way ANOVA to test for differences in activity between the 3 orientations. There was no significant difference in activity between orientations for either WM or IM in any of the ROIs (all $p > 0.30$).

2.7. Calculating % signal change

To convert the activity time courses to % signal change we used the following conversion:

$$\%signal\ change = \left(\frac{task\ activity}{baseline} \times 100\right) - 100$$

As baseline we took the average activity in a voxel at the first scan of each trial, over all WM and IM trials in that run together. To test whether the difference in BOLD signal between preferred and non-preferred voxels was reliable, we used one-sample *t*-tests against zero.

2.8. Relationship between overall BOLD activity, decoding and behavior

To investigate whether there was a relationship between decoding accuracy and BOLD differences between preferred and non-preferred orientations during the maintenance of stimulus representations, we first calculated Spearman correlation coefficients over participants between the BOLD difference and decoding accuracy for WM and IM in the period 8–10 s after onset of the imagery period. We chose these time points to be maximally distant in time from activity induced by the two stimuli presented at the start of each trial, but early enough to avoid contamination with activity related to the probe, which was presented at 10 s.

To test whether overall BOLD related to decoding, we also calculated Spearman correlation coefficients over participants between decoding accuracy at 8–10 s with overall BOLD level at 8–10 s. Subsequently, to quantify the similarity between the average decoding accuracy time course and the average BOLD difference time course for each region, we calculated Spearman rank correlations between the two time courses for each participant and averaged these. Moreover, we calculated a Spearman rank correlation between the two average time courses.

We also correlated decoding accuracy with overall BOLD activity during perception at 4 s after stimulus presentation. Furthermore, we directly compared the magnitude of the activity difference for preferred and non-preferred orientation during stimulus visualization just before the probe (at 8–10 s) with the magnitude of the activity difference after visual presentation of the probe (at 14 s) using paired *t*-tests, per region and for WM and IM independently. Finally, we similarly calculated the correlation between the BOLD differences and behavioral performance (angular difference of the probe and imagined orientation).

2.9. Separating orientations based on their 'role' in the task

To test whether a BOLD difference between preferred and non-preferred stimulus representations also occurred for orientations that had not been visually presented, we investigated trials in which people had to imagine an orientation that was not presented in that particular trial. For those trials, we further separated the time course difference based on the different 'roles' of the orientations during a trial: (i) the orientation of the stimulus that was presented and cued, (ii) the orientation that was presented, but neither cued, nor the target of mental rotation and (iii) the target orientation of the mental rotation process, as those also showed distinct decoding time courses (Albers et al., 2013).

2.10. Relationship between activity differences and decoding performance

To investigate whether the difference between preferred and non-preferred stimulus representations was due to a higher response to preferred orientations or a decreased response to non-preferred orientations, we performed a median split on the participants based on the average of how well we could decode the different orientations for WM and IM (high decoding accuracy and low decoding accuracy, $N = 12$ per group). We then examined the overall BOLD signal in preferred and non-preferred voxels for these two groups separately. Specifically, we performed a 3-way (task by preference by group) ANOVA to test whether activity for preferred and non-preferred orientations differed per task and per group.

2.11. Regional percentages

To quantify the division of activity over central and peripheral early visual cortex, we used the ROIs obtained using the templates from Benson and colleagues (Benson et al., 2012, 2014). Within these larger masks, we selected all the voxels with a *t*-value > 0 and obtained their eccentricity values. Subsequently, we calculated the average eccentricity over all included voxels and performed, per ROI, paired *t*-tests to compare the eccentricity of activity during *the visual stimulation contrast* with the *task probe contrast* and the *task stimulus contrast.* As a control, we also performed these analyses using the median and with the nonparametric Wilcoxon signed-rank test, which provided similar results.

3. Results

3.1. Small differences in activity between a voxel's preferred and non-preferred orientation

We investigated a previously reported discrepancy between decoding accuracy and overall BOLD levels (e.g. Albers et al., 2013; Harrison and Tong, 2009). To this end, voxels in visual cortex were split into groups with different stimulus-orientation preferences ('sub-ROIs'), and the time course of the BOLD response to preferred and non-preferred orientations was extracted. We found that the average BOLD amplitude time courses for preferred and non-preferred stimuli largely resembled each other (WM: Fig. 2A; IM: Fig. 2B). However, there was a numerically small, yet

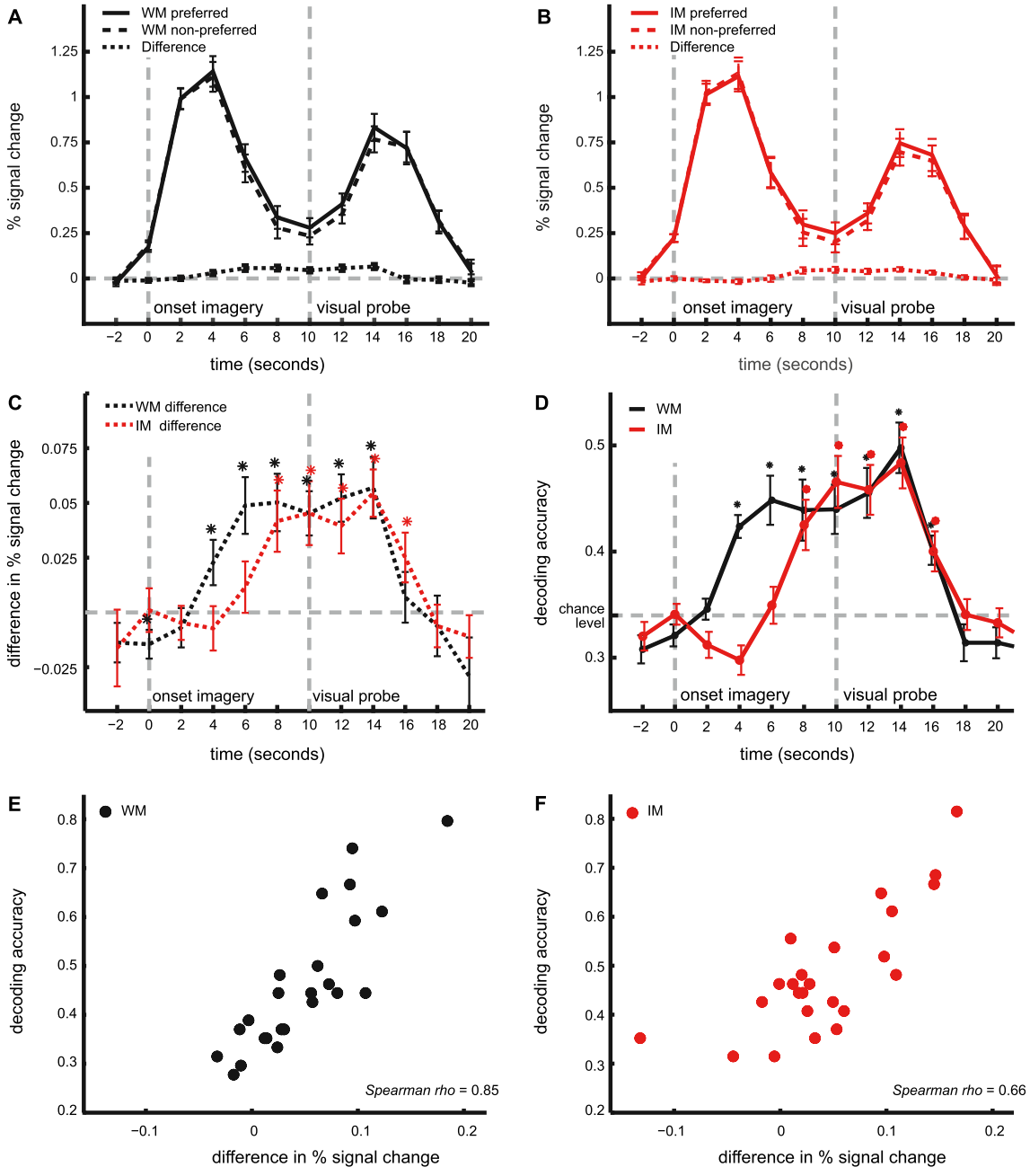

Fig. 2. Time course of BOLD signal change and decoding in early visual cortex (V1-V3 combined). A) Time course of mean BOLD activity (averaged over the 360 selected voxels from central V1-V3) during working memory (WM) trials, separately for trials in which the preferred orientation was the target (solid lines) and trials in which the non-preferred orientation was the target (dashed lines). For both preferred and non-preferred orientations the neural activity peaked 4–6 s after onset of the stimulus presentation and again 4–6 s after presentation of the probe. In between, the activity decreased substantially. The time course of activity for preferred and non-preferred orientations was highly similar: the difference between the two hardly differed from zero (dotted line just above baseline). Percent signal change was calculated with respect to the scans immediately preceding each trial (time point −2 s). Error bars denote SEM; the horizontal grey dashed line indicates baseline; vertical grey lines indicate timing of events in the trial. **B)** Time course of mean neural activity during mental imagery (IM) trials, other conventions as in A). **C)** Magnification of the time course of the differences in activity between preferred and non-preferred orientations (dotted lines in A) and B)). The shape of the time course resembles the shape of the decoding time course in D). Asterisks indicate a difference that was significantly different from 0, which occurred at 4–16 s for WM trials and at 8–16 s for IM trials (* indicates $p < 0.05$, two-tailed). **D)** Time course of decoding accuracy for working memory (black line) and mental imagery (red line), reproduced with permission from Albers et al. (2013). Significant decoding (indicated by asterisks) of the target orientation occurred at 4–16 s for WM and at 8–16 s for IM trials (*all* $p < 0.001$), the same time points as in C). The delayed decoding ability for imagined versus maintained stimulus representations is expected from the task. **E)** Correlation between BOLD difference and decoding accuracy for WM trials at 8–10s. There was a strong relationship (WM: Spearman ρ (22) = 0.85, $p < 0.00001$) between inter-individual differences in decoding accuracy and differences in BOLD activity between preferred and non-preferred orientations. Black dots represent single subjects. **F)** Correlation between BOLD difference and decoding accuracy for IM trials, as in E). Also for IM there was a strong correlation between inter-individual differences in decoding accuracy and BOLD activity difference (IM: Spearman ρ (22) = 0.66, $p = 0.0002$).

consistent *difference* in BOLD activity to preferred orientation and to non-preferred orientations (Fig. 2C). One-sample t-tests on these differences, over subjects, showed significant differences at 4–14 s (WM) and 8–16 s (IM) after onset of the task (*all* $p < 0.05$, two-tailed; * in Fig. 2C).

The time course of the *difference* between preferred and non-preferred orientations was very similar to the classification performance time course. When correlating the *average time courses*, all correlations were above 0.9 (V1-V3 WM: *Spearman* $\rho = 0.95$, $p < 0.0001$; V1-V3 IM: $\rho = 0.97$, $p < 0.0001$; for correlations per ROI see Table 1), when taking the average of correlations per participant the correlations were somewhat lower, but significantly larger than 0 (V1-V3 WM: *Spearman* $\rho = 0.56$, $p < 0.0001$; V1-V3 IM: $\rho = 0.51$, $p < 0.0001$; for correlations per ROI see Table 1). Significant decoding (Fig. 2D, * indicate $p < 0.001$) occurred at the time points at which also a significant difference in BOLD was observed. Moreover, between-subject variance in BOLD activity differences at 8–10 s strongly correlated with variance in decoding accuracy in early visual cortex at that same point in time (WM: *Spearman* $\rho = 0.85$, $p < 0.0001$; IM: $\rho = 0.66$, $p < 0.0001$; Fig. 2E–F). This difference between activity for preferred versus non-preferred orientations, and the correlation with decoding accuracy was also present in V1, V2 and V3 independently (Table 1 and Supplementary Fig. S1; the somewhat smaller decoding accuracy and correlation with the BOLD difference in V1 are in line with earlier findings (e.g. Bosch et al., 2014; Harrison and Tong, 2009; Kamitani and Tong, 2005)).

Since decoding accuracy, in turn, was related to better performance on the comparison task (Albers et al., 2013) we also examined the correlation between behavioral accuracy and between-subject variance in the BOLD activity difference. A larger BOLD difference for WM trials indeed correlated with better performance (a smaller angle difference) on the WM task (*Spearman* $\rho = -0.48$, $p = 0.01$); for IM this effect was in the same direction but not significant (*Spearman* $\rho = -0.24$, $p = 0.13$).

Further divisions of the time course showed that the BOLD effect was independent of visual perception: there was also a higher BOLD response to preferred orientations when they were imagined but had not been visually presented. That is, on some trials people imagined an orientation that was not presented on that particular trial and those orientations that were merely imagined also led to an increase in BOLD signal specifically in the voxels that preferred the imagined orientation (Fig. 3A); again, the time course of the differences resembled the time course of decoding (Fig. 3B; (Albers et al., 2013)). Together, these findings suggest that the SVM classification relied on populations of voxels that were more active during trials with their preferred orientation than during trials with their non-preferred orientation(s), but were largely obscured by the overall BOLD modulations.

3.2. Differences in decoding relate to a larger response to preferred orientations

Is the larger difference between preferred and non-preferred orientations in participants with better decoding due to an increased response to preferred stimuli, or is it the effect of a weaker response (or a stronger suppression) to a stimulus that they do not prefer? To investigate this issue, we split the participants in two groups using a median split (N = 12 per group), based on their decoding accuracy. We used an ANOVA to test for group (high or low decoding accuracy), preference (preferred or non-preferred orientation) and task (WM or IM) effects and interactions between them, on the level of BOLD activity during the delay period (8–10 s).

Overall activity during the delay period tended to be higher for the group with higher decoding (Fig. 4; main effect of group ($F(1, 23) = 3.69$, $p = 0.068$); see Fig. S2 for decoding plots split over participants) and although preferred orientations led to higher activity than non-preferred orientations for both groups (main effect of preference, $F(1, 23) = 4.29$, $p < 0.001$), this effect was larger for participants that showed better decoding (group by preference interaction, $F(2, 21) = 16.3$, $p < 0.001$). These effects did not differ for the two tasks (WM and IM). We next directly compared activity for preferred and non-preferred stimulus representations between the two groups. These *post hoc* tests showed that activity to preferred orientations (averaged over WM and IM) was higher for participants with high decoding accuracy (0.40% signal change) than for participants with low decoding accuracy (0.15% signal change; $F(1, 22) = 4.78$, $p = 0.04$). On the other hand, there was no significant difference in activity for non-preferred orientations between participants with high decoding accuracy (0.32% signal change) and participants with low decoding accuracy (0.13% signal change; $F(1, 22) = 2.7$, $p = 0.11$). This suggests that the difference between preferred and non-preferred is due to higher activity to preferred orientations. Finally, we directly correlated overall BOLD activity and decoding accuracy over participants using Spearman rank correlation, but found no significant correlation between overall level of BOLD during maintenance (8–10 s) and decoding accuracy at that same time point for either WM (*Spearman* $\rho = 0.26$, $p = 0.11$), or IM (*Spearman* $\rho = 0.01$, $p = 0.49$).

3.3. A stimulus-unspecific response in peripheral early visual cortex

The differences in activity between preferred and non-preferred voxels were much smaller than the overall BOLD response during maintenance/imagery and visual stimulation (Fig. 2A & B vs C). At the peak of activity in response to the probe (14 s), which coincided with the peak of the difference and the peak of the decoding, the activity difference between preferred and non-preferred stimulus representations was only 8% (WM) and 7% (IM) of the overall BOLD response. Interestingly, the size of the orientation-specific difference during the visual presentation of the probe (at 14 s) did not differ from the size of the orientation-specific difference during imagery (at 8–10 s; *all* $p > 0.16$), during which the overall BOLD response was markedly lower. In other words, the bulk of the BOLD signal to the probe stimulus did not appear to be "orientation-specific".

Previous studies that reported a non-stimulus related BOLD response in early visual cortex during a decision process suggested that this signal was largely located in peripheral regions of early visual cortex (Donner et al., 2008; Jack et al., 2006). To investigate whether the overall BOLD response during the decision process in this case also contained such a non-specific component, we investigated its spatial topography, as we

Table 1
Correlation between the activity difference and decoding accuracy. Spearman rank correlations over participants between decoding accuracy and the difference in activity to preferred and non-preferred orientations at time points 8–10, when taking the average of the correlations between time courses for each participant, and when correlating the *average time courses*, for V1, V2 and V3 separately and combined.

	V1	V2	V3	V1-V3
8-10 s WM	$\rho(22) = 0.48$, $p = 0.0082$	$\rho(22) = 0.80$, $p < 0.00001$	$\rho(22) = 0.83$, $p < 0.00001$	$\rho(22) = 0.85$, $p < 0.00001$
8-10 s IM	$\rho(22) = 0.59$, $p = 0.0013$	$\rho(22) = 0.66$, $p = 0.0002$	$\rho(22) = 0.77$, $p < 0.00001$	$\rho(22) = 0.66$, $p = 0.0002$
Average time course correlations per participant WM	$\rho(22) = 0.39$, $p < 0.0001$	$\rho(22) = 0.56$, $p < 0.0001$	$\rho(22) = 0.58$, $p < 0.0001$	$\rho(22) = 0.56$, $p < 0.0001$
Average time course correlations per participant IM	$\rho(22) = 0.38$, $p < 0.0001$	$\rho(22) = 0.47$, $p < 0.0001$	$\rho(22) = 0.48$, $p < 0.0001$	$\rho(22) = 0.51$, $p < 0.0001$
Correlation over average time courses WM	$\rho = 0.8462$, $p = 0.0010$	$\rho = 0.9441$, $p < 0.0001$	$\rho = 0.9231$, $p < 0.0001$	$\rho = 0.9510$, $p < 0.0001$
Correlation over average time courses IM	$\rho = 0.7622$, $p = 0.0059$	$\rho = 0.9301$, $p < 0.0001$	$\rho = 0.9371$, $p < 0.0001$	$\rho = 0.9650$, $p < 0.0001$

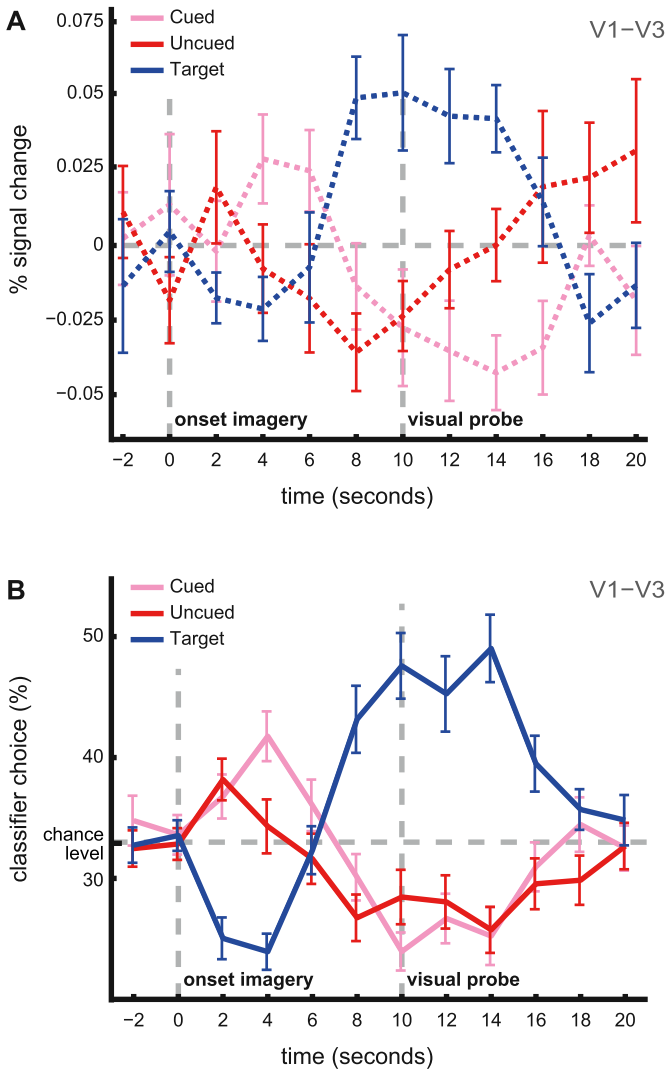

Fig. 3. Temporal unfolding of the BOLD difference time course and decoding accuracy for the different stimulus representations in V1-V3. A) Differences in neural activity between preferred and non-preferred orientations (as in Fig. 2C), for trials in which participants mentally rotated the cued stimulus toward the not-physically presented grating. Time courses are plotted separately for the different 'roles' of the orientations during the task: presented and cued stimulus (pink), presented but uncued stimulus (red) and target orientation after rotation (blue). During the first time points, there was a difference in activity between preferred and non-preferred orientations for the two physically presented stimuli (red and pink lines), but not for the not-presented grating. Thereafter, there was a gradual switch of the larger difference in activity from the cued grating (pink) to the generated target grating (blue). This shows that the larger activity for preferred compared to non-preferred orientations was also present when participants imagined a stimulus that they had not previously perceived (blue line). Error bars denote SEM; the grey dashed line indicates baseline; vertical dashed grey lines indicate timing of events in the trial. Percent signal change was calculated over the 360 selected voxels in V1-V3 with respect to the scans immediately preceding each trial (time point −2 s). **B)** Temporal pattern of decoding for the cued, uncued and imagined (target) gratings, for trials in which participants mentally rotated the cued stimulus toward the not-physically presented grating; reproduced with permission from Albers et al. (2013). Decoding indicated by the proportion of classifier choice when testing V1-V3 combined, averaged over the 24 participants. The time course of BOLD differences between preferred and non-preferred orientations in A) can also explain these decoding results. Error bars denote SEM; horizontal dashed grey line indicates chance level (33.3%); vertical grey dashed lines indicate events in the trial.

expected such a stimulus-unspecific signal to appear in peripheral regions as well. Specifically, we compared the activity during the presentation of the probe (*task probe contrast*), during which participants made a decision, with the statistical activity maps during bottom-up stimulus processing obtained during the localizer (*visual stimulation contrast*) and with stimulus processing during the main task (*task stimulus contrast*). To quantify the spread over central and peripheral regions, we selected all voxels with *t*-values > 0 in the V1—V3 ROIs based on the templates from Benson et al. (2012, 2014), and quantified their eccentricity (Supplementary Fig. S3). We subsequently tested whether activity occurred at larger eccentricities for the *task probe contrast* than for the *visual stimulation contrast* and the *task stimulus contrast* (Table 2).

Stimulus processing (i.e., the stimulus presentation during the localizer; *visual stimulation contrast*) resulted in strong, focal activity in central early visual cortex, the regions that were covered by the stimulus (average eccentricity 8.28°; Supplementary Fig. S3). The activity during the stimulus and cue presentation in the main task (*task stimulus contrast*, first peak in the time course) also clustered in central regions, but extended more into peripheral eccentricities (average eccentricity 15.32° (±2.42). The neural activity during the probe (*task probe contrast*, second peak in the time course) spread out along all eccentricities (average eccentricity 17.21° (±1.36)). Accordingly, the average eccentricity of activated voxels during *visual stimulation contrast* was much smaller than the average eccentricity of the activated voxels during the *task stimulus contrast* ($t(23) = -10.77$, $p = 2.1 \times 10^{-9}$), and also smaller than the eccentricity during the *task probe contrast* ($t(23) = -14.98$, $p = 2.1 \times 10^{-12}$). The spread of activity during the *task probe contrast* was larger than during the *task stimulus contrast* ($t(23) = -4.45$, $p = 0.0017$; all *p*-values corrected for multiple comparisons). Comparable effects were observed in V2 and V3 (Table 2). This suggests that there is indeed a large component of non-stimulus-specific activity in early visual cortex during probe presentation, as well as, to a smaller extent, during stimulus presentation.

4. Discussion

Here we investigated how reliable stimulus information can be decoded from patterns of BOLD activity in early visual cortex during mental maintenance of stimulus representations when BOLD amplitude is low or even at baseline levels (e.g. Bosch et al., 2014; Emrich et al., 2013; Harrison and Tong, 2009). We reanalyzed a previously reported dataset in which participants mentally imagined or maintained in working memory an oriented grating (Albers et al., 2013). Although the overall BOLD activity in early visual cortex was highly stereotypical and independent of stimulus content, there was a slight but consistent increase in BOLD signal in those voxels that were tuned to the imagined orientation, compared to voxels that preferred another stimulus orientation. The size of these differences strongly correlated with the decoding accuracy, as well as with the participants' behavioral performance. The overall BOLD increase during visual presentation of the probe contained both a stimulus-specific and a large stimulus-unspecific component, covering both central (stimulus-covered) and peripheral (non-stimulated) regions of V1-V3.

The current results suggest that the stimulus-evoked BOLD signals in visual cortex during a cognitive task are an amalgam of stimulus-specific and non-specific processes, whereby a substantial part of the BOLD response to stimuli may be non-specific, both in terms of orientation selectivity and retinotopic specificity, hiding small stimulus-specific modulations. However, this study indicates that, on top of this large non-specific contribution, a small but reliable difference in stimulus-specific activity can be found. These findings may link the earlier findings of univariate decreases in activity but multivariate increases in information during working memory maintenance (e.g. Albers et al., 2013; Harrison and Tong, 2009; Sneve et al., 2012). Namely, a decodable stimulus representation can result from a highly specific pattern of activity in that subset of voxels relevant for encoding the stimulus, which is (partly) obscured by large non-specific modulations. The matching time courses for decoding and BOLD differences also suggest that it is this stimulus-specific BOLD component that underlies the ability to read out stimulus information from activity patterns (Albers et al., 2013; Bosch et al., 2014; Christophel et al., 2015; Harrison and Tong, 2009), whereas

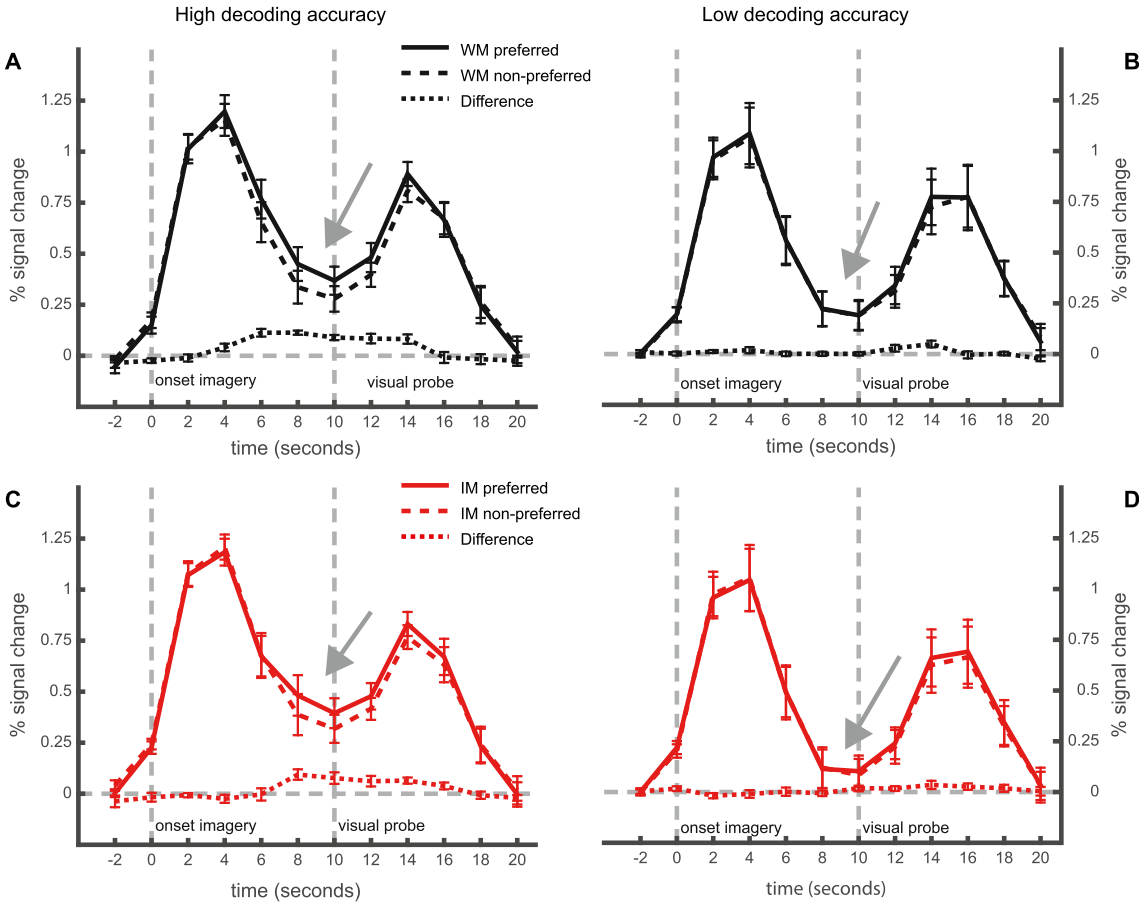

Fig. 4. The BOLD activity for participants with high vs. low decoding accuracy. A) Time course of WM activity (over 360 selected voxels from V1-V3) for preferred (solid lines) and non-preferred (dashed lines) orientations, for the group of participants where we found high decoding accuracy (N = 12; median split on decoding accuracy WM and IM combined); dotted line just above the baseline (at zero): the time course of the difference in activity for preferred and non-preferred orientations. Percent signal change was calculated with respect to the scans immediately preceding each trial (time point −2 s). Error bars denote SEM; the horizontal grey dashed line indicates baseline; vertical grey lines indicate timing of events in the trial. **B)** Time course of WM activity for preferred (solid lines) and non-preferred (dashed lines) orientations, for the group of participants with low decoding accuracy (N = 12); as in A). **C)** Time course of IM activity for preferred (solid lines) and non-preferred (dashed lines) orientations, for the group of participants where we found high decoding accuracy (N = 12; median split); as in A). **D)** Time course of IM activity for preferred (solid lines) and non-preferred (dashed lines) orientations, for the group of participants with low decoding accuracy (N = 12); as in A).

Table 2
Average eccentricities per ROI and comparison of average eccentricities during visual stimulation and presentation of stimuli and probe. Average eccentricity per contrast and ROI calculated over all the voxels within the full V1, V2 and V3 ROIs with *t*-values > 0. Eccentricity values were obtained from the templates provided by Benson and colleagues (Benson et al., 2012, 2014). The contrasts were compared per ROI, using paired *t*-tests. All *p*-values are corrected for the multiple comparisons.

	V1	V2	V3
Average eccentricity per ROI			
Visual stimulation contrast	8.28 (±2.52)	7.21 (±1.93)	6.79 (±1.63)
Task stimulus contrasts	15.32 (±2.42)	13.42 (±3.41)	9.98 (±2.12)
Task probe Contrast	17.21 (±1.36)	16.65 (±1.69)	13.05 (±2.03)
Comparison of average eccentricities for the different contrasts			
Visual stimulation contrast vs. *task stimulus contrast*	$t\,(23) = -10.77$, $p = 1.7 \times 10^{-9}$	$t\,(23) = -8.10$, $p = 3.1 \times 10^{-7}$	$t\,(23) = -6.33$, $p = 1.7 \times 10^{-5}$
Visual stimulation contrast vs. *task probe contrast*	$t\,(23) = -14.98$, $p = 2.1 \times 10^{-12}$	$t\,(23) = -19.84$, $p = 5.2 \times 10^{-15}$	$t\,(23) = -13.55$, $p = 1.7 \times 10^{-11}$
Task stimulus contrasts vs. *task probe contrast*	$t\,(23) = -4.45$, $p = 0.0017$	$t\,(23) = -5.82$, $p = 5.6 \times 10^{-5}$	$t\,(23) = -6.36$, $p = 1.6 \times 10^{-5}$

the overall BOLD response was unrelated to the stimulus information.

4.1. A univariate signal that underlies multivariate results

The current findings fit with the presence of a specific BOLD signal in relevant voxels, partly obscured by a large, unspecific BOLD modulation. However, they also indicate that representations of orientation do not necessarily rely on a multivariate code at the voxel level that is independent of overall activity levels. A multivariate code – an intricate pattern of interactions between the activity in several voxels, where the stimulus cannot be read out from the activity in one or several voxels but only when looking at activation and deactivation in multiple voxels combined – is sometimes implicitly assumed to underlie the results obtained with multivariate pattern analysis. The current findings suggest an underlying univariate effect at the voxel level, whereby the activity in one or multiple subpopulation(s) of voxels is indicative of the stimulus orientation. That univariate differences can underlie the results of a multivariate analysis was one of the original ideas behind decoding (Norman et al., 2006; Tong et al., 2012) and has been reiterated by theoretical and simulation studies (Anderson and Oates, 2010; Davis et al., 2014; Haufe et al., 2014; Mur et al., 2009) and our data provide empirical evidence for this notion.

What is considered 'univariate' or 'multivariate' might depend on the scale at which one investigates the brain: single voxels can code for a stimulus in a 'univariate' manner, as we observe in this study, while when considering the entire cortical region V1, multiple voxels may be

needed to read out which of multiple stimuli has been seen or imagined, resembling a multivariate code at the level of ROIs. Here we showed how very small univariate effects at the voxel level can explain a multivariate finding in early visual cortex. The sensitivity of MVPA decoding to such small, spatially distributed and sparse (univariate) effects (Davis et al., 2014; Jimura and Poldrack, 2012; Mur et al., 2009), is one of the features that makes MVPA a powerful analysis tool for investigating the presence of (inherently) subtle signals related to mental representations. This does obviously not imply that all decoding results rely on such small univariate differences as we find here; large overall stimulus-related BOLD changes, or a multivariate effect combining information from several voxels might underlie several findings obtained with MVPA analyses. Also in our dataset there are participants with high decoding accuracy but small differences between preferred and non-preferred stimuli. Nevertheless, in this study the bulk of the experimental variance emerges as a univariate effect at the voxel level. Encoding models that explicitly model how the BOLD patterns encode certain features (e.g. Brouwer and Heeger, 2009; Davis et al., 2014; Mitchell et al., 2008; Naselaris et al., 2015; Vu et al., 2011) might eventually provide a more fruitful approach to understanding the (univariate or multivariate) nature of the underlying neural code.

4.2. Increased signal specificity for maintained stimuli

The small, univariate increase in response to preferred orientations over non-preferred orientations indicate that the signal becomes more specific during maintenance of internally generated material. The link between the size of the effect and behavioral accuracy also support earlier findings that link activity patterns in V1 to working memory ability (Emrich et al., 2013; Ester et al., 2013). It seems relevant to know the nature of the feedback signals that influence those internally generated stimulus representations. Do the small differences in activity between preferred and non-preferred stimulus representations originate from a stronger response to preferred orientations, from a weaker response (or larger suppression) to non-preferred orientations, or from a combination of these effects? We observed that activity for both preferred and non-preferred stimulus representations was stronger in those participants with better decoding. Furthermore, the response to preferred orientations showed a stronger boost than the response to non-preferred orientations in the group of participants with high decoding accuracy. This suggests that stronger activity for preferred stimuli, rather than inhibition of activity to non-preferred stimuli, is a major contributor to the decodability of the representations. Interestingly, this effect did not differ for WM and IM, even though on half of the IM trials the orientation had not been visually presented. The stronger response to preferred orientations suggests that there is top-down *activation* of relevant neurons for internally generated content.

4.3. Large, non-specific BOLD modulations during decisions

During presentation of the probe, the large, overall BOLD signal modulation was not restricted to stimulus-specific areas in visual cortex, but spread out to peripheral parts of early visual cortex as well. This supports the idea that part of the BOLD activation in this region was not related to stimulus perception, in line with earlier studies that found similar effects during task transitions and decision-related processing (Donner et al., 2008; Jack et al., 2006; Kleinschmidt, 2006; Kloosterman et al., 2015; Swallow et al., 2012; Warren et al., 2015). Indeed, the signal appeared when the participants had to judge the probe and give a button press response. Together, this suggests that the overall BOLD signal during our task consisted of both stimulus-specific and non-stimulus-specific signals.

A second factor that might contribute to the small magnitude of the stimulus-specific differences is that the stimulus-unspecific activity could also result from voxels that are equally responsive to all three orientations and therefore did not show differential activity. As activity in a voxel is pooled from a large number of underlying neurons with different orientation tuning, each voxel is likely to respond to all orientations at least to some extent. This might especially hold for the voxels in the central regions of early visual cortex, which were selected based on their response amplitude for all gratings alike. A highly similar overall response with small differences in orientation preference in different voxels, is in line with the original rationale for orientation decoding in early visual cortex (Kamitani and Tong, 2005). However, this notion does not account for the fact that stimulus-specific differences were of equal magnitude during maintenance and probe presentation, while overall BOLD signals were vastly different.

4.4. Relationship between BOLD activity and decoding

Finally, our results speak to the relationship between overall BOLD activity and decoding accuracy. Orientation-specific BOLD differences were largest during the delay period, when the overall BOLD signal was relatively low; the relevant signals thus appeared to be independent of the overall BOLD modulation. Yet, the average BOLD activity to preferred orientations at 8–10 s was higher in the group of participants where we found better decoding. This suggests that decoding accuracy might be coupled to BOLD amplitude during mental imagery of oriented gratings – even though there is no physical input when stimulus representations are internally generated (Tong et al., 2012). Potentially, there is no qualitative difference with earlier findings that the classification performance of *perceived* gratings was highly dependent on the amplitude of the stimulus-driven activity (Smith et al., 2011; Tong et al., 2012). The explanation of these earlier findings, with decoding being dependent on the SNR of the *differential activity* (Tong et al., 2012), also matches with our findings: a more distinct signal for different orientations that underlies the better decoding accuracy. Such an increased response to preferred stimuli could enhance overall activity as well. One might speculate that, since BOLD and decoding relate to stimulus strength in the case of perception, in case of internally generated representations higher BOLD and decoding indicate a stronger internal stimulus representation as well. Future experiments could test such a claim by relating BOLD and decoding accuracy on a trial by trial level.

5. Conclusions

Our results suggest that stimulus-evoked BOLD signals in visual cortex during a task context are an amalgam of stimulus-specific and non-stimulus-specific processes, whereby a large part of the BOLD response to stimuli can be non-specific, both in terms of orientation selectivity and retinotopic specificity. On top of this relatively large orientation-unselective contribution is a small but reliable difference in stimulus-specific activity. This latter activity correlated highly with the decoding accuracy and constituted the likely source of stimulus decoding. Together, these effects may explain why decoding of stimulus orientation (which relies on stimulus-specific BOLD signal differences) can be highly accurate even when overall BOLD signals are low.

Acknowledgements

The authors would like to thank Peter Kok for help with data collection and initial decoding analyses. Funding for this study was provided by the Netherlands Organisation for Scientific Research (NWO Brain & Cognition 433-09-248).

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.neuroimage.2017.09.046.

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. Curr. Biol. 23, 1427–1431. https://doi.org/10.1016/j.cub.2013.05.065.

Anderson, M.L., Oates, T., 2010. A critique of multi-voxel pattern analysis. In: Proceedings of the 32nd Annual Meeting of the Cognitive Science Society. Cognitive Science Society, pp. 1511–1516.

Benson, N.C., Butt, O.H., Brainard, D.H., Aguirre, G.K., 2014. Correction of distortion in flattened representations of the cortical surface allows prediction of V1-V3 functional organization from anatomy. PLoS Comput. Biol. 10, e1003538 https://doi.org/10.1371/journal.pcbi.1003538.

Benson, N.C., Butt, O.H., Datta, R., Radoeva, P.D., Brainard, D.H., Aguirre, G.K., 2012. The retinotopic organization of striate cortex is well predicted by surface topology. Curr. Biol. 22, 2081–2085. https://doi.org/10.1016/j.cub.2012.09.014.

Bosch, S.E., Jehee, J.F., Fernandez, G., Doeller, C.F., 2014. Reinstatement of associative memories in early visual cortex is signaled by the hippocampus. J. Neurosci. 34, 7493–7500. https://doi.org/10.1523/JNEUROSCI.0805-14.2014.

Brouwer, G.J., Heeger, D.J., 2009. Decoding and reconstructing color from responses in human visual cortex. J. Neurosci. 29, 13992–14003. https://doi.org/10.1523/JNEUROSCI.3577-09.2009.

Christophel, T.B., Cichy, R.M., Hebart, M.N., Haynes, J.D., 2015. Parietal and early visual cortices encode working memory content across mental transformations. NeuroImage 106, 198–206. https://doi.org/10.1016/j.neuroimage.2014.11.018.

Cukur, T., Nishimoto, S., Huth, A.G., Gallant, J.L., 2013. Attention during natural vision warps semantic representation across the human brain. Nat. Neurosci. 16, 763–770. https://doi.org/10.1038/nn.3381.

Davis, T., LaRocque, K.F., Mumford, J.A., Norman, K.A., Wagner, A.D., Poldrack, R.A., 2014. What do differences between multi-voxel and univariate analysis mean? How subject-, voxel-, and trial-level variance impact fMRI analysis. NeuroImage 97, 271–283. https://doi.org/10.1016/j.neuroimage.2014.04.037.

DeYoe, E.A., Carman, G.J., Bandettini, P., Glickman, S., Wieser, J., Cox, R., Miller, D., Neitz, J., 1996. Mapping striate and extrastriate visual areas in human cerebral cortex. Proc. Natl. Acad. Sci. U. S. A. 93, 2382–2386. https://doi.org/10.1073/pnas.93.6.2382.

Donner, T.H., Sagi, D., Bonneh, Y.S., Heeger, D.J., 2008. Opposite neural signatures of motion-induced blindness in human dorsal and ventral visual cortex. J. Neurosci. 28, 10298–10310. https://doi.org/10.1523/JNEUROSCI.2371-08.2008.

Emrich, S.M., Riggall, A.C., Larocque, J.J., Postle, B.R., 2013. Distributed patterns of activity in sensory cortex reflect the precision of multiple items maintained in visual short-term memory. J. Neurosci. 33, 6516–6523. https://doi.org/10.1523/JNEUROSCI.5732-12.2013.

Engel, S.A., Glover, G.H., Wandell, B.A., 1997. Retinotopic organization in human visual cortex and the spatial precision of functional MRI. Cereb. Cortex 7, 181–192. https://doi.org/10.1093/cercor/7.2.181.

Ester, E.F., Anderson, D.E., Serences, J.T., Awh, E., 2013. A neural measure of precision in visual working memory. J. Cogn. Neurosci. 25, 754–761. https://doi.org/10.1162/jocn_a_00357.

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

Haufe, S., Meinecke, F., Gorgen, K., Dahne, S., Haynes, J.D., Blankertz, B., Biessmann, F., 2014. On the interpretation of weight vectors of linear models in multivariate neuroimaging. NeuroImage 87, 96–110. https://doi.org/10.1016/j.neuroimage.2013.10.067.

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

Horikawa, T., Tamaki, M., Miyawaki, Y., Kamitani, Y., 2013. Neural decoding of visual imagery during sleep. Science 340, 639–642. https://doi.org/10.1126/science.1234330.

Jack, A.I., Shulman, G.L., Snyder, A.Z., McAvoy, M., Corbetta, M., 2006. Separate modulations of human V1 associated with spatial attention and task structure. Neuron 51, 135–147. https://doi.org/10.1016/j.neuron.2006.06.003.

Jimura, K., Poldrack, R.A., 2012. Analyses of regional-average activation and multivoxel pattern information tell complementary stories. Neuropsychologia 50, 544–552. https://doi.org/10.1016/j.neuropsychologia.2011.11.007.

Kamitani, Y., Tong, F., 2005. Decoding the visual and subjective contents of the human brain. Nat. Neurosci. 8, 679–685. https://doi.org/10.1038/nn1444.

Kamitani, Y., Tong, F., 2006. Decoding seen and attended motion directions from activity in the human visual cortex. Curr. Biol. 16, 1096–1102. https://doi.org/10.1016/j.cub.2006.04.003.

Kleinschmidt, A., 2006. Cognitive control signals in visual cortex: flashes meet spotlights. Neuron 51, 9–11. https://doi.org/10.1016/j.neuron.2006.06.010.

Kloosterman, N.A., Meindertsma, T., Hillebrand, A., van Dijk, B.W., Lamme, V.A., Donner, T.H., 2015. Top-down modulation in human visual cortex predicts the stability of a perceptual illusion. J. Neurophysiol. 113, 1063–1076. https://doi.org/10.1152/jn.00338.2014.

Kok, P., Jehee, J.F., de Lange, F.P., 2012. Less is more: expectation sharpens representations in the primary visual cortex. Neuron 75, 265–270. https://doi.org/10.1016/j.neuron.2012.04.034.

Kriegeskorte, N., Goebel, R., Bandettini, P., 2006. Information-based functional brain mapping. Proc. Natl. Acad. Sci. U. S. A. 103, 3863–3868. https://doi.org/10.1073/pnas.0600244103.

Mitchell, T.M., Shinkareva, S.V., Carlson, A., Chang, K.M., Malave, V.L., Mason, R.A., Just, M.A., 2008. Predicting human brain activity associated with the meanings of nouns. Science 320, 1191–1195. https://doi.org/10.1126/science.1152876.

Mur, M., Bandettini, P.A., Kriegeskorte, N., 2009. Revealing representational content with pattern-information fMRI–an introductory guide. Soc. Cogn. Affect Neurosci. 4, 101–109. https://doi.org/10.1093/scan/nsn044.

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.

Norman, K.A., Polyn, S.M., Detre, G.J., Haxby, J.V., 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn. Sci. 10, 424–430. https://doi.org/10.1016/j.tics.2006.07.005. S1364–6613(06)00184-7 [pii].

Riggall, A.C., Postle, B.R., 2012. The relationship between working memory storage and elevated activity as measured with functional magnetic resonance imaging. J. Neurosci. 32, 12990–12998. https://doi.org/10.1523/JNEUROSCI.1892-12.2012.

Serences, J.T., Ester, E.F., Vogel, E.K., Awh, E., 2009. Stimulus-specific delay activity in human primary visual cortex. Psychol. Sci. 20, 207–214. https://doi.org/10.1111/j.1467-9280.2009.02276.x.

Sereno, M.I., Dale, A.M., Reppas, J.B., Kwong, K.K., Belliveau, J.W., Brady, T.J., Rosen, B.R., Tootell, R.B., 1995. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science 268, 889–893. https://doi.org/10.1126/science.7754376.

Smith, A.T., Kosillo, P., Williams, A.L., 2011. The confounding effect of response amplitude on MVPA performance measures. NeuroImage 56, 525–530. https://doi.org/10.1016/j.neuroimage.2010.05.079.

Sneve, M.H., Alnaes, D., Endestad, T., Greenlee, M.W., Magnussen, S., 2012. Visual short-term memory: activity supporting encoding and maintenance in retinotopic visual cortex. NeuroImage 63, 166–178. https://doi.org/10.1016/j.neuroimage.2012.06.053.

Swallow, K.M., Makovski, T., Jiang, Y.V., 2012. Selection of events in time enhances activity throughout early visual cortex. J. Neurophysiol. 108, 3239–3252. https://doi.org/10.1152/jn.00472.2012.

Tong, F., Harrison, S.A., Dewey, J.A., Kamitani, Y., 2012. Relationship between BOLD amplitude and pattern classification of orientation-selective activity in the human visual cortex. NeuroImage 63, 1212–1222. https://doi.org/10.1016/j.neuroimage.2012.08.005.

Vu, V.Q., Ravikumar, P., Naselaris, T., Kay, K.N., Gallant, J.L., Yu, B., 2011. Encoding and decoding v1 fmri responses to natural images with sparse nonparametric models. Ann. Appl. Stat. 5, 1159–1182. https://doi.org/10.1214/11-AOAS476.

Wandell, B.A., Dumoulin, S.O., Brewer, A.A., 2007. Visual field maps in human cortex. Neuron 56, 366–383. https://doi.org/10.1016/j.neuron.2007.10.012.

Warren, C.M., Nieuwenhuis, S., Donner, T.H., 2015. Perceptual choice boosts network stability: effect of neuromodulation? Trends Cogn. Sci. 19, 362–364. https://doi.org/10.1016/j.tics.2015.05.007.

Woolgar, A., Golland, P., Bode, S., 2014. Coping with confounds in multivoxel pattern analysis: what should we do about reaction time differences? A comment on Todd, Nystrom & Cohen 2013. NeuroImage 98, 506–512.