
Access from the University of Nottingham repository:
http://eprints.nottingham.ac.uk/2641/1/Denis_Schluppeck--decoding_working.pdf

Copyright and reuse:

The Nottingham ePrints service makes this work by researchers of the University of Nottingham available open access under the following conditions.

This article is made available under the Creative Commons Attribution Non-commercial Share Alike licence and may be reused according to the conditions of the licence. For more details see: http://creativecommons.org/licenses/by-nc-sa/2.5/

A note on versions:

The version presented here may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the repository url above for details on accessing the published version and note that access may require a subscription.

For more information, please contact eprints@nottingham.ac.uk
Behavioral/Cognitive

Decoding Working Memory of Stimulus Contrast in Early Visual Cortex

Yue Xing, Tim Ledgeway, Paul V. McGraw, and Denis Schluppeck
School of Psychology, University of Nottingham, Nottingham NG7 2RD, United Kingdom

Most studies of the early stages of visual analysis (V1-V3) have focused on the properties of neurons that support processing of elemental features of a visual stimulus or scene, such as local contrast, orientation, or direction of motion. Recent evidence from electrophysiology and neuroimaging studies, however, suggests that early visual cortex may also play a role in retaining stimulus representations in memory for short periods. For example, fMRI responses obtained during the delay period between two presentations of an oriented visual stimulus can be used to decode the remembered stimulus orientation with multivariate pattern analysis. Here, we investigated whether orientation is a special case or if this phenomenon generalizes to working memory traces of other visual features. We found that multivariate classification of fMRI signals from human visual cortex could be used to decode the contrast of a perceived stimulus even when the mean response changes were accounted for, suggesting some consistent spatial signal for contrast in these areas. Strikingly, we found that fMRI responses also supported decoding of contrast when the stimulus had to be remembered. Furthermore, classification generalized from perceived to remembered stimuli and vice versa, implying that the corresponding pattern of responses in early visual cortex were highly consistent. In additional analyses, we show that stimulus decoding here is driven by biases depending on stimulus eccentricity. This places important constraints on the interpretation for decoding stimulus properties for which cortical processing is known to vary with eccentricity, such as contrast, color, spatial frequency, and temporal frequency.

Introduction
Several lines of research suggest that the human brain has a specific cognitive system for holding information to be manipulated and executed: working memory. Visual short-term memory (VSTM) is a specific subtype that allows the robust maintenance of stimulus attributes such as contrast, orientation, spatial frequency, and speed with high fidelity (Luck and Vogel, 1997; Pasternak and Greenlee, 2005). Functional brain imaging has enabled the exploration of the anatomical and functional correlates underlying VSTM first discovered by neurophysiological techniques (Fuster, 1995; Braver et al., 1997; Pessoa et al., 2002; Curtis and D’Esposito, 2003). Although studies largely agree on the involvement of higher-level cortical areas in this cognitive process (Postle and D’Esposito, 1999; Haxby et al., 2000), there is still some controversy about the role of early visual cortex in working memory. Recent fMRI evidence suggests that early sensory areas may be involved in retaining stimulus representations, for example, of orientation (Pessoa et al., 2002; Serences et al., 2009) and spatial frequency (Greenlee et al., 2000; Baumann et al., 2008; Sneve et al., 2011). However, the sustained responses in early visual cortex may be related to visual attention rather than VSTM (Offen et al., 2009, 2010; Pooresmaeili et al., 2010). Harrison and Tong (2009) used multivariate pattern analysis (MVPA) to search for potential signatures of working memory for orientation in the pattern of fMRI responses in early visual areas. They found that remembered orientation could be decoded and that the same neural circuitry that mediates early visual processing (and perception) of orientation is also recruited during the working memory period.

The aim of our experiments was to determine whether a different stimulus property—the contrast of a stimulus—could be decoded using multivariate classification and if its representation was similar when subjects perceived and remembered stimuli. Most neurons in visual cortex show tight tuning for a preferred stimulus property, such as local contrast, orientation, or direction of motion. Recent evidence from electrophysiology and neuroimaging studies, however, suggests that early visual cortex may be related to visual attention rather than VSTM (Offen et al., 2009, 2010; Pooresmaeili et al., 2010). Harrison and Tong (2009) used multivariate pattern analysis (MVPA) to search for potential signatures of working memory for orientation in the pattern of fMRI responses in early visual areas. They found that remembered orientation could be decoded and that the same neural circuitry that mediates early visual processing (and perception) of orientation is also recruited during the working memory period.

The first experiment in this study shows that the stimulus contrast, as well as the orientation, of a perceived image can be decoded from event-related BOLD signals in early visual cortex. The second experiment demonstrates that the contrast of a remembered grating (during periods when no stimulus is displayed) can be also decoded from the BOLD responses in visual cortex.
We further show that classifiers generalize between the two experiments and that classification accuracies were significantly higher for behaviorally correct than incorrect trials, indicating that signals from early visual cortex contribute significantly to VSTM of stimulus contrast. Our results also suggest that responses from incorrect trials add substantial noise to the contrast VSTM signals used in decoding.

Materials and Methods

Subjects. Six observers (n = 5 male, n = 1 female) who were experienced in psychophysics and fMRI experiments and had normal or corrected-to-normal vision took part in this study. All gave written consent. The procedures were approved by the Medical School Research Ethics Committee at the University of Nottingham.

Functional MRI acquisition. Each subject participated in at least two scanning sessions. In session one, a set of 10–12 functional scans was obtained to measure the retinotopic organization in early visual cortex, allowing us to functionally define V1, V2, and V3 with standard methods; in the same session, we also acquired high-resolution anatomical T1-weighted MPRAGE images of the whole brain for segmentation and cortical flattening. In the second (and later) sessions, we obtained fMRI data to perform decoding in the working memory and stimulus contrast paradigms.

MR imaging was performed at 3 T (Achieva; Philips Healthcare) using an eight-channel SENSE head coil. Foam padding was used to minimize head movements. During each session, we acquired several functional scans, including a scan for localizing regions of early visual cortex representing the stimuli and six to eight scans for the main tasks, alternating between scans in which we measured responses to perceived stimuli and scans using the working memory task. For BOLD fMRI, we used a standard T2* (gradient-echo) echo planar imaging pulse sequence (voxel size 3 × 3 × 3 mm3, TE = 35 ms, TR = 1500 ms, flip angle = 75°, FOV = 192 × 192 mm2). Thirty-two slices were oriented approximately perpendicularly to the calcarine sulcus. We used parallel imaging with an acceleration (SENSE) factor 2.

Registration, cortical segmentation, and flattening. For flat mapping and visualization, we segmented T1-weighted anatomical images into gray matter, white matter, and CSF. Inflated and flattened parts of each hemisphere corresponding to early visual cortex were created using a combination of tools (FreeSurfer, Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School, Boston, MA; mrTools, Heeger lab, Department of Psychology and Center for Neuroscience, New York University, New York, NY; mrVISTA, Wandell lab, Department of Psychology, Stanford University, Stanford, CA) and programs included in the FSL distribution (FMRIB Software Library; Smith et al., 2004).

To register data from each session to the subject-specific, high-resolution, T1-weighted image, a set of low-resolution anatomical images covering the same volume as the functional images was acquired at either the beginning or the end of each scanning session (MPRAGE, 1.5 mm inplane, 3 mm slice thickness). These anatomical images and the functional images were then registered to the high-resolution anatomical whole-head volume (T1-weighted, 3D MPRAGE, 1 mm isotropic, TE = 3.7 ms, TR = 8.13 ms, FA = 8°, TI = 960 ms, and linear phase encoding order) using a robust image registration algorithm (Nestares and Heeger, 2000).

Visual stimuli. In all experiments, the stimuli were generated on an Apple MacBook Pro running MATLAB (MathWorks) and the MGL toolbox (http://gru.brain.riken.jp/doku.php/mlg/overview). In the fMRI experiments, stimuli on a homogenous gray background were projected from an LCD projector onto a display screen at the feet of our subjects. The display resolution was 1024 × 768 pixels, covering 20.4° (width) × 15.4° (height) of visual angle. Subjects were in the supine position in the scanner bore and viewed the display through an angled mirror. They were asked to maintain fixation at a red cross at the center of the screen during functional MRI scans and performed a task in all scans to control attention (see below). All of the stimuli in the two experiments were moving sinusoidal gratings (spatial frequency, 0.75 cycles/°; temporal frequency, 2 Hz) presented in a circular aperture with a radius of 5°, centered at fixation.

Retinotopic mapping session. Early visual areas (V1, V2, and V3) for each subject were identified in a retinotopic mapping session based on the standard traveling-wave method using rotating wedges and expanding rings (Engel et al., 1994; DeYoe et al., 1996; Engel et al., 1997; for review, see Wandell et al., 2007). The responses to the rings and wedges were used to estimate the eccentricity and polar angle of the visual field representation, respectively. Following standard methods, areas V1, V2, and V3 were defined in our subjects using the phase reversals in the polar angle maps to locate the upper, lower, and horizontal meridian representations.

Localizer scans. During each functional imaging session of the main experiment, we obtained two brief localizer scans (at the beginning and end of each session) to identify voxels within V1-V3 corresponding to the retinotopic stimulus locations. We used these to restrict the V1-V3 regions of interest (ROIs). Each localizer scan lasted 4 min (160 time points) and consisted of grating stimuli and gray background alternating in a simple block design between 12 s "on" and 12 s "off" (fixation). During the on period, we presented 8.1 s epochs of moving gratings each with a random direction of motion and at 1 ° of contrast levels (Michelson contrasts of 10%, 30%, 50%, 70%, and 90%). For gamma correction, we used a psychophysical motion-nulling procedure (Lefèvre and Legge, 1994; Lefèvre and Hess, 2002) to make sure that contrast stimuli would be faithfully reproduced on the LCD projector display at the MRI scanner and on a CRT display in the psychophysics laboratory.

Experiment 1: contrast response measurements. To measure the responses in early visual cortex to the stimuli, we presented moving gratings (orientation 45° or 135°, spatial frequency, 0.75 cycles/°; temporal frequency, 2 Hz, same 5 contrast levels used in the localizer scan) in an event-related design. Stimuli were presented in random order (preceded by 1.5 s of fixation to equate for conditions in the working memory scans). Each stimulus lasted 1.5 s and subjects were asked to respond as quickly as possible to their offset by pressing a button to maintain a constant attentional state.

Experiment 2: working memory scans. Consistent with the experimental paradigm used by Harrison and Tong (2009), we used an event-related design in which each trial comprised visual stimuli presentation, a retention period, and a test interval (Fig. 1). The same stimulus parameters were used as before except that there were only two levels of contrast in this experiment. During the stimulus presentation at the beginning of each trial, two gratings with different contrasts (30% and 70%) were presented sequentially. Each grating presentation (in a circular aperture of radius 5° centered at fixation) was accompanied by a small text label ("1" or "2") rendered in red and displayed with a small offset from the fixation cross; the assignment of the label to the higher or lower contrast grating was randomized on each trial. The orientation of both gratings in each trial was chosen to be either 45° or 135°, but remained constant throughout the trial. Briefly, we estimated—voxelwise—the numeric cue indicated which of the two contrast stimuli the subjects should remember and retain in memory for the delay period (fixed, 8.4 s). Finally, at the end of each trial, a matching grating (contrast ± 20% of the to-be-remembered grating) was displayed and the subjects were asked to indicate with a press of a button whether the remembered (sample) or matching grating had the higher contrast.

fMRI response time courses. Imaging data were analyzed using a combination of custom-written software (mrTools, Heeger lab, Department of Psychology and Center for Neuroscience, New York University, New York, NY; mrVISTA, Department of Psychology, Stanford University, Stanford, CA) running in MATLAB 7.4. fMRI data were motion corrected using a robust motion correction algorithm (Nestares and Heeger, 2000), high-pass filtered (cutoff, 0.01 Hz) to remove slow signal drift, and converted to percent signal change.

Standard event-related analysis methods (Burock et al., 1998; Dale, 1999; Burock and Dale, 2000) were used to reconstruct the time course of fMRI response for each trial type at each voxel with no predefined assumption of the hemodynamic response function. Details of the methodology have been reported in detail previously (Gardner et al., 2005; Brouwer and Heeger, 2009; Sanchez-Panchuelo et al., 2010). Briefly, we estimated—voxelwise—the best-fitting (least-squares) event-related time course for each stimulus condition. To estimate the goodness-of-fit of the linear model prediction, we
calculated the amount of variance explained in the original time course for each voxel by our model \( r^2 \). The average fMRI time series was then computed over all voxels within an ROI (Fig. 2A) that met a threshold \( r^2 \) of 0.2. This average time series (calculated from a fixed number of voxels per ROI) was then used for further analyses.

For Experiment 1 (contrast response measurements), we used the maximum values (±1 time point) of the average event-related fMRI response for each trial type (Fig. 2B, inset) to plot contrast-response curves to estimate the increase in fMRI responses in each ROI with stimulus contrast. For Experiment 2 (the VSTM-for-contrast experiment), we ascertained that there were no systematic response amplitude differences in the average event-related fMRI response between the two trial types in the ROIs because this could be a potential confound for the multivariate classification analysis (Fig. 2C).

MVPA. MVPA was used to analyze data from both experiments. Classification was performed with a linear support vector machine (SVM) implemented in custom-written MATLAB functions.

Each measurement (trial) of fMRI response for a different stimulus condition (class) at different voxels corresponds to a point in a high-dimensional pattern space. We obtained each point (which could also be considered a spatial pattern of response) by averaging several time points of the fMRI responses in a single trial taking into account the sluggishness of the hemodynamic response. For the remembered-contrast experiment, we also took care to avoid possible temporal overlap of the BOLD response during the memory period with that evoked by the second (comparison) interval in each trial. For the contrast response measurements, data were sampled and averaged across three time points (TRs) centered on the peak response in each trial to account for with variability in the hemodynamic delay for each subject. For the working memory task, the time points corresponding to the retention period after the stimulus presentation (TRs 5–7, which were 7.5–10.5 s after stimulus onset) for each subject were selected and averaged to maximize the coverage of the memory period (Harrison and Tong, 2009). Therefore, an \( m \times n \) matrix served as the input for the classification analysis, where \( m \) is the number of trials and \( n \) is the number of voxels included in the analysis.

A linear SVM provides optimal parameter values (weights and bias) for a hyperplane to categorize two classes of data (collected under two different stimulus conditions) within such an N-dimensional space. To access how generalizable the classification was, we tested our analysis with a 10-fold (or 5-fold) cross-validation scheme with the two classes approximately matched in each fold: 90% (80%) of data were used to train a classifier and the remaining, hold-out 10% (20%) of data to test. To maximize the number of samples, we did not average trials before classification. We report the mean classification accuracy across these 10 (5) folds.

For classification of the contrast response data (which had five stimulus classes), we trained the model on pairs of stimulus classes using part of the data and then used it to predict one of the two contrast levels with the rest of data. The best-predicted label for each class was chosen by a winner-take-all rule. Trials in which the stimulus type was mislabeled were considered as errors in this particular analysis.

The statistical significance of the observed classification accuracies for each subject was assessed by a binomial test, considering the probability of correctly labeled test examples of all the independent testing data points is given by the binomial distribution \( p < 0.05 \); null hypothesis: stimulus class cannot be predicted because the classifier was performing at chance, 50%). Statistical significance across subjects was assessed with a \( t \) test (two-tailed, \( p < 0.05 \)).

**Results**

**Contrast-response functions**

We first performed a control experiment to determine how reliable the information in the human early visual areas was for decoding the contrast of visually presented stimuli in a task that did not involve a VSTM component. The responses to stimuli with different contrasts were measured and the time courses for these different stimulus conditions (averaged across all voxels in V1, V2, and V3 that met a threshold \( r^2 \) of 0.2, respectively) were examined. Figure 2B shows the response time courses in V1 of one subject for gratings at five different contrast levels with a systematic increase in the overall fMRI response with increasing contrast. The same pattern of response was found across all the other observers. We obtained contrast-response functions (Fig. 2B, inset; curve fit: standard Naka-Rushton function) by averaging the peak response amplitude for each contrast level in a time window around the peak response (±1 time point). The aim of this control experiment was primarily to confirm that, for the
stimuli we used, the measured fMRI responses were not at floor or ceiling levels.

**Experiment 1: multivariate classification of stimulus contrast**

To establish whether the fMRI responses in V1-V3 contain consistent information about the contrast of the stimuli beyond that conveyed in the mean time course across all voxels in the ROI, we studied the pattern of responses using multivariate classification analysis. Classification accuracies in V1 and V2 ROIs were all significantly above chance across our group of subjects (Fig. 3A).

To ensure that the pattern of results in the classification accuracies obtained did not depend critically on our choice of voxel number, we performed the following analysis. We randomly selected a range of 2–210 voxels (in increments of 15) from the 210 candidate voxels in each V1, V2, and V3 and performed the classification analysis with this subset of features. We then repeated this analysis in 100 bootstrapped replications for each subject and plotted classification accuracy as a function of the number of voxels included. The inset in Figure 3A shows the mean (±1 SEM) for data in V1 across six subjects, indicating that there were small differences in classification accuracy that depend on voxel number but that, importantly, the overall pattern of results remained consistent. (The dashed line indicates the classification accuracy for n = 180 voxels, the lowest common number of voxels available across ROIs and subjects rounded down to a multiple of 10).

We expected that correctly classifying trials with large contrast differences should be easier than for small contrast differences. To test this, we performed a finer grained analysis of the pairwise classifications that contributed to the five-way classification. In four separate analyses, we split trials into groups according to their differences in linear contrast (Δ = 0.2, 0.4, 0.6, 0.8). Note that the number of trials that could be included in the analysis therefore differed, ranging from all trials for the pairwise analysis of trials with small contrast differences (Δ = 0.2) to only the lowest and highest contrast trials for the pairwise analysis with a contrast difference of 0.8. The pattern of prediction accuracies for V1 is shown in Figure 3B. Classification accuracies between trials increased with the difference in contrast between the grating patterns regardless of the stimulus orien-

---

**Figure 2.** Retinotopic mapping and event-related responses in early visual cortex. **A**, Example retinotopic maps from one subject. Information from visual angle and eccentricity measurements were used to define V1, V2 (ventral + dorsal), and V3 (ventral + dorsal). **B**, Event-related fMRI responses as a function of visual stimulus contrast in V1 of one subject. Gray lines indicate the estimated event-related responses for low (light-gray) to high (dark gray) contrast. Error bars indicate SEM across trials. Inset: fMRI contrast-response curves in V1 (red), V2 (green), and V3 (blue) for this subject; symbols, amplitude of measured fMRI response; lines, nonlinear least-squares fit to data (Naka-Rushton). **C**, Event-related responses from the working memory experiment. Plot shows average responses in V1 during trials in which the remembered contrasts were 0.3 and 0.7, respectively (mean ± SEM across subjects). The curves show a typical pattern of response with a large transient after the presentation of the two stimuli and cue at fixation at the beginning of each trial, followed by a reduced response during the delay period and another transient increase in response after the presentation of the comparison stimulus.
tation. V2 and V3 showed the same dependence on the difference in stimulus contrast.

Because the classification analysis could potentially have been driven mainly by differences in the mean response level across different stimulus conditions, we also performed a control analysis in which we classified trials only using the mean response level across the ROI. Figure 3 illustrates how we calculated the receiver operating characteristic (ROC) and measured the area under curve for each of the possible pairwise comparisons. We constructed histograms of the mean fMRI response in the ROI for two classes of trials, which were expected to have a different mean across trials. For example, the top panel in Figure 3C shows the distribution of response amplitudes for trials with 10% contrast (class 1, red line) and 90% contrast (class 2, gray curves). Follow-
ing the logic of a standard signal-detection theory analysis, we set a criterion level and estimated the proportion of trials from the two distributions that fell above that criterion level. This corresponds to the proportion of “hits” (or correct labelings) and “false alarms” (or incorrect labelings), respectively. Considering the proportion of hits and false alarms at different criterion levels, we could construct the ROC curve. The area under the ROC curve is equivalent to the proportion correct that could be achieved in two-way categorization given the distribution of fMRI responses for the two classes of stimuli (Macmillan and Creelman, 2005).

Interestingly, although the classification performance improved as the contrast difference between the two gratings increased, those obtained by only considering the mean response were consistently worse than those obtained by multivariate classification. This indicates that there is significant categorical information present in the pattern of responses that is not captured in the ROI mean, which can be exploited by using multivariate methods.

The scatter plot in Figure 3D summarizes the data for the four different levels of contrast difference and four different ROIs. For all pairwise comparisons of trials with a given difference in contrast (size of symbols) in all of the ROIs we considered (color of symbols), multivariate classification outperformed classification based solely on the mean response in the ROI.

**Experiment 2: multivariate classification of remembered contrast**

In the second part of the study, we addressed whether BOLD responses in early visual cortex could be used to classify the remembered stimulus contrast of a pattern. We recorded the BOLD responses when two levels of contrasts (“low,” 30% or “high,” 70%) were remembered. Average time courses in V1 for the two types of trials across six subjects showed no apparent differences between the responses during trials when subjects had to remember the low or high contrast pattern (Fig. 2C). All subjects’ responses across V1, V2, and V3 in this task showed a similar profile.

The first transient increase in the event-related responses follows the presentation of the sample gratings and cue at the beginning of each trial. The second transient response corresponds to the presentation of the test grating. The period in between corresponds to the interval during which subjects had to remember the contrast of the grating to allow them to perform the task. We averaged the four time points in this time window (see Materials and Methods) and used these data as input for our classification analysis.

Using multivariate classification, we could successfully categorize the remembered contrast trials with accuracies significantly above chance level, as shown in Figure 4A. This result was consistent across data from early visual areas V1 and V2. Results for V3 showed a similar trend, but failed to reach significance.

**Analysis of correct/incorrect trials**

We also split the fMRI data according to behavioral performance on each trial and considered classification accuracies separately for correct and incorrect trials, respectively. This analysis revealed that decoding remembered contrast levels from early visual cortex was significantly better when subjects had correctly remembered the stimuli (Fig. 4B, n = 4 subjects for whom there was a sufficient number of incorrect trials to perform the analysis). This suggests that the pattern of fMRI responses was more consistent and reproducible across “correct” trials than “incorrect” trials, indicating an increased consistency of response patterns that leads to higher classification accuracies. The behavioral performance of our subjects in the working memory task for contrast indicated that the participants performed the task reliably. There was no significant difference in performance for maintaining in memory a pattern with a contrast of 30% compared with a grating with a contrast of 70% (74.1% correct vs 73.8% correct; t(90) = 0.471), implying that task difficulty was nearly matched across the two conditions. This may indicate that errors are more related to memory lapses or the inability to respond in time rather than difficulties with encoding the contrast level during the presentation of the sample pattern.

**Generalization analysis**

To further examine the role of early visual areas in maintaining a working memory representation for contrast, we tried to predict the remembered contrast during the memory task (30% and 70% contrast) using the SVM classifier trained with the data collected during (control) Experiment 1. If the response patterns of perceived stimuli could be generalized to those of remembered stimuli successfully, this could provide additional evidence to implicate the function of early visual cortex in short-term memory for contrast.

To match the amount of data used in training and test of classification, the following 10-fold cross-validation procedure was used: 90% of data samples from Experiment 1 were used to train an SVM classifier, whose ability to classify was tested with 10% of data from Experiment 2 (working memory). Classification accuracies in this analysis were above chance for V1 and V2, with a similar trend for V3 (Fig. 4C).

In the converse analysis, training on memory trials and then testing on data from perceived stimuli, performance also exceeded chance level for V1. We also found that generalization from perceived to remembered stimuli was most consistent when the contrast for training and test data were matched (Fig. 4D, black bars) and were reduced to chance for increasingly dissimilar training stimuli. These results provide compelling evidence that the patterns of response in early visual cortex associated with subjects performing the VSTM task and perceiving stimuli are highly similar, although it does not imply a causal role of these responses in visual working memory.

**Spatial pattern of responses underlying contrast decoding**

To characterize the spatial pattern of responses that underlie decoding in our experiment, we subdivided the visual areas into two sets of ROIs: one set based on the eccentricity measurements from retinotopic mapping into a central (0.5°–2.5° eccentricity) and an eccentric ROI (2.5°–5°). Figure 5 shows results from V1. As a control, we also defined a second set of ROIs corresponding to the upper and lower visual field, which subdivided V1 orthogonally to the central/eccentric split. We compared the responses in these ROIs in two ways. First, we plotted the contrast-response functions for the two sets of ROIs (Fig. 5C,D) and, second, we used the mean time course from the two respective ROIs to create two supervoxels (Freeman et al., 2011), which could then be used in the classification analyses.

The contrast-response curves are clearly different for the two splits (Fig. 5C,D), showing a consistent response difference for central/eccentric but not the upper/lower visual fields. (Note that increasing the number of supervoxels to four had no appreciable effect on the pattern of results as seen in Fig. 6). Next, we used the responses in the two sets of supervoxels to show that the differences at this large scale can drive classification of perceived contrast, even when only two features are considered. If the supervoxels are constructed from a central/eccentric split, classification is possible; if, however, data are reduced to two supervoxels corresponding to the upper and lower visual fields, classification accuracies drop to chance (Fig. 5E).
The corresponding analysis for data from the visual working memory experiment shows that an eccentricity bias also drives the classification in the working memory experiment (Fig. 5F), suggesting that inhomogeneity in the spatial pattern of visual cortex responses—potentially driven by a feedback signal or attention—underlies the memory "decoding."

Discussion
A wealth of studies have confirmed that multivariate classification can be used to decode basic features of visual stimuli and even to predict attended stimulus features, in particular the orientation of gratings (Haynes and Rees, 2005; Kamitani and Tong, 2005, 2006; Mannion et al., 2009; Sapountzis et al., 2010; Swisher...
Spatial pattern of fMRI responses that drive classification. **A, B**, Flat map of visual cortex from one subject showing definition of central/eccentric and lower/upper visual field ROIs for V1. White dashed lines indicate borders between V1 and V2v/V2d. **C, D**, FMRI response as a function of stimulus contrast. Contrast response curves from the central/eccentric (Figure legend continues.)

**Figure 5.** Spatial pattern of fMRI responses that drive classification. **A, B**, Flat map of visual cortex from one subject showing definition of central/eccentric and lower/upper visual field ROIs for V1. White dashed lines indicate borders between V1 and V2v/V2d. **C, D**, FMRI response as a function of stimulus contrast. Contrast response curves from the central/eccentric (Figure legend continues.)
et al., 2010). However, there is substantial debate concerning the source of the underlying signal supporting this decoding both for stimulus orientation (Freeman et al., 2011) and other stimulus parameters (Kriegeskorte, 2009).

Here we show that fMRI responses in early visual cortex can also be used to reliably decode the stimulus contrasts presented to the subject. The ability to classify trials with different contrast levels is itself unsurprising because there is a well known shift in mean response level with increasing contrast, which has previously been well characterized in early visual cortex (Albrecht and Hamilton, 1982; Ohzawa et al., 1985; Boynton et al., 1999; Reich et al., 2001; Gardner et al., 2005). Most neurons in early visual cortex exhibit a monotonic increase in response as a function of contrast, but some cells show very different patterns of response saturation (and even supersaturation) with contrast (Ledgeway et al., 2005; Peirce, 2007; Hu and Wang, 2011). However, unlike stimulus orientation, there is no known columnar organization for neurons with similar contrast-response functions. Nonetheless, despite the lack of orderly clustering in structure, there may still be spatially local inhomogeneous changes in the fMRI responses to contrast. Using a multivariate classification analysis that does not take into account shifts in the mean response in the ROI therefore cannot untangle whether the ability to classify is simply due to a global shift in the neuronal response or to redistribution of a differential signal.

In a more fine-grained analysis, we therefore evaluated multivariate classification between stimuli at pairwise contrast levels and the predicted performance using a univariate (signal detection type) analysis using only the average response in each ROI. We found that multivariate classification became increasingly accurate with larger contrast differences in the stimuli (Fig. 3B; Smith et al., 2011). Even trials close in contrast could be discriminated using SVM, where just considering the mean response in the ROI failed to classify trials correctly. This further supports MVPA’s increased sensitivity compared with univariate methods (Kriegeskorte et al., 2006).

These results suggest that a spatially specific response pattern to different contrasts can drive improved classification performance. To characterize the spatial pattern (and scale) of responses underlying this multivariate contrast decoding, we subdivided the visual areas into two sets of ROIs: one based on eccentricity measurements from retinotopic mapping corresponding to a central and an eccentric portion of our stimuli, because contrast sensitivity is known to change as a function of eccentricity (Legge and Kersten, 1987). As a control, we also defined another set of ROIs corresponding to the upper and lower visual fields, where such asymmetries might be less pronounced. We found that even when we reduced the multivariate fMRI responses to only two features, responses in a central and eccentric ROI, classification was possible. Our results do not, however, address directly whether information is present only at the level of a coarser scale (eccentricity) map or also at fine columnar architecture, an issue currently the subject of much debate (Gardner et al., 2008; Op de Beeck et al., 2008; Gardner, 2010; Swisher et al., 2010; Freeman et al., 2011; Beckett et al., 2012).

To determine whether MVPA could also be used to probe the mechanisms underlying working memory of stimulus contrast, we analyzed fMRI responses during a delay period in a visual working memory task. Given previous findings (Ester et al., 2009; Harrison and Tong, 2009) that a significant working memory trace exists for specific orientations in early visual cortex, we investigated whether a similar result holds for contrast.

We found that information in the fMRI responses during working memory trials could be used to decode whether a low-contrast or high-contrast visual pattern was retained in memory. Interestingly, there was no significant difference in mean response in early visual cortex when subjects remembered a low-contrast or high-contrast visual pattern was retained in memory. It should be noted that the experimental design used here (consistent with Harrison and Tong’s paradigm) is not ideally suited to estimate sustained delay period activity because the delay duration is fixed (see Schluppeck et al., 2006 for discussion) and by necessity always follows a cor directly. Exact estimation of delay period activity in such a paradigm is therefore ill posed. However, to the extent that sustained delay period activity can be estimated, there was variability across our six subjects in the amount of sustained response, which is in agreement with previous reports (Harrison and Tong, 2009).
contrast versus a high-contrast stimulus, suggesting that overall response level is not the representation used for remembering contrast. The pattern of fMRI responses in early visual cortex, however, contained sufficient information to decode trials into the two categories. This result is consistent with Offen et al. (2009), who proposed an explanation for the absence of (mean) delay period activity in early visual areas during working memory tasks: opposing excitatory and inhibitory responses from different subpopulations of neurons in these regions may effectively cancel each other in the average BOLD signal. Any delay period activity in early visual cortex may therefore be obscured in individual voxels or the mean responses across ROIs.

To address the key issue of whether the pattern of fMRI responses obtained when subjects perceived a given stimulus contrast is similar to when they remembered it, we assessed the ability of our classifiers to generalize across the different tasks (Kamitani and Tong, 2005, 2006; Dinstein et al., 2008; Kay et al., 2008; Brouwer and Heeger, 2009; Harrison and Tong, 2009). In Harrison and Tong’s (2009) study, they predicted which of two orientations were retained in memory with a classifier trained on data from visual activity patterns induced by unattended gratings. In the present study, generalization was tested by comparing classification with perceived versus remembered stimulus contrasts. We found that classification generalized both from perceived to remembered and vice versa, strongly suggesting that the representation of information in early visual cortex (specifically V1) is similar for both tasks.

Intriguingly, when we reduced the multivariate responses into those of two supervoxels corresponding to a central and an eccentric portion of our stimuli, classification of remembered contrasts was still possible. This suggests that a large-scale bias with eccentricity can drive classification for remembered as well as perceived stimuli.

Moreover, we observed a strong dependence of classification performance on the behavioral performance of our subjects, which is consistent with previous reports (Williams et al., 2007; Scolari and Serences, 2010). When trials were split according to correct and incorrect behavioral performance, we found that classification accuracies for correct trials were significantly higher than for incorrect trials. This suggests that the pattern of responses in early visual cortex is more consistent and repeatable across correct trials than incorrect trials. If multivariate classification analysis relies on signals directly related to neural activity supporting working memory function, then this increased consistency in fMRI responses may correspond to decreased noise and a more reliable neural signal. Activity in early visual cortex during incorrect trials may be more inconsistent, because there are several different reasons for making errors, causing more variability across trials: lapses in attention, eye blinks, failure to encode or retain the matching stimulus, and finger (response) errors.

Some studies suggest that both task-relevant and task-irrelevant features may be encoded together in VSTM (O’Craven et al., 1999; Wheeler and Treisman, 2002), whereas others have argued for a task-selective activity pattern in primary visual cortex (Woodman and Vogel, 2008; Serences et al., 2009). One previous fMRI experiment used color or orientation as selected features to evaluate prediction accuracy of MVPA in a working memory task (Serences et al., 2009). That study found a significant change of performance when subjects’ attention was switched between the alternate features in different runs. Here, we did not test directly whether the information from two orientations that were task irrelevant were retained in memory (subjects only judged stimulus contrast), but, like others (Serences et al., 2009), we used two primary stimulus attributes (orientation and stimulus contrast) to reduce adaptation effects across trials.

Our study shows that fMRI responses in early visual cortex could be used to decode the contrast of a perceived stimulus (Experiment 1) and that responses in these areas also supported the classification of contrast when the stimulus only had to be remembered (Experiment 2). Classifiers trained on data from each experiment generalized to the other, suggesting that signals in early visual cortex are significantly modulated during working memory for stimulus contrast on the timescale of seconds and that the same signals are present during perception and memory for this stimulus property. We found that a large-scale bias in the responses with eccentricity can drive classification for remembered as well as perceived stimuli, raising the possibility that a consistent attentional or feedback signal, rather than activity related to working memory per se, may underlie the significant (but modest) classification accuracies reported.

Finally, we found substantially improved pattern classification when we compared data from correct and incorrect trials. When we considered only data from correct trials for classification, accuracies approximately matched subjects’ behavioral performance. This highlights the fact that fMRI responses from incorrect trials add substantial noise to the contrast VSTM signals used in decoding, an important consideration for future studies.

References
Ester EF, Serences JT, Awh E (2009). Spatially global representations in hu-