nature
neuroscience

Decoding the visual and subjective contents of the human brain

Yukiyasu Kamitani[1] & Frank Tong[2,3]

The potential for human neuroimaging to read out the detailed contents of a person's mental state has yet to be fully explored. We investigated whether the perception of edge orientation, a fundamental visual feature, can be decoded from human brain activity measured with functional magnetic resonance imaging (fMRI). Using statistical algorithms to classify brain states, we found that ensemble fMRI signals in early visual areas could reliably predict on individual trials which of eight stimulus orientations the subject was seeing. Moreover, when subjects had to attend to one of two overlapping orthogonal gratings, feature-based attention strongly biased ensemble activity toward the attended orientation. These results demonstrate that fMRI activity patterns in early visual areas, including primary visual cortex (V1), contain detailed orientation information that can reliably predict subjective perception. Our approach provides a framework for the readout of fine-tuned representations in the human brain and their subjective contents.

Much remains to be learned about how the human brain represents basic attributes of visual experience. It is commonly assumed that human visual perception is based on the neural coding of fundamental features, such as orientation, color, motion and so forth. Edge orientation is arguably the most fundamental feature analyzed by the visual system, providing the basis for defining the contours and shapes of visual objects that allow for object segmentation and recognition. There is considerable neurophysiological evidence of orientation tuning in cortical columns and single neurons in both monkey and cat visual cortex[1–4]. However, non-invasive neuroimaging methods have been thought to lack the resolution to probe into these feature representations in the human brain. As a consequence, little is known about the neural basis of orientation selectivity in the human visual system or how these fundamental features are represented during conscious perception.

Here we investigated whether it is possible to read out detailed information about the visual and subjective contents of the human brain using fMRI. Specifically, can activity patterns in the human visual cortex reveal what stimulus orientation a person is viewing or attending to on a given trial? We present a new approach to measure feature selectivity from an ensemble of neuroimaging signals, which we call 'ensemble feature selectivity'. Our approach draws on ideas of population coding[5,6] and multi-voxel pattern analysis of fMRI data[7–9] to demonstrate neural decoding of perceived orientations and a method of 'mind-reading' that is capable of classifying mental states on the basis of measured brain states. Early versions of this work have been presented at scientific conferences (Y.K. and F.T., *J. Vis.* **4**, 186a, 2004; Y.K. and F.T., *Soc. Neurosci. Abstr.* 370.7, 2004).

A challenge in examining orientation selectivity in the human visual cortex is that putative orientation columns may be too finely spaced to resolve using current fMRI techniques. Orientation-selective columns in the monkey are only about 300–500 mm in width[10], whereas the spatial resolution of human fMRI is limited by many factors. These include technical limitations of human fMRI, reductions in signal-to-noise proportional to the volume of each voxel, spatial blurring of the positive blood oxygenation level-dependent (BOLD) hemodynamic response extending several millimeters beyond the site of neural activity[11–13], and additional blurring induced by residual head motion.

To bypass these spatial limitations, we developed an alternative approach of measuring the ensemble orientation information contained in the activity pattern of many voxels. We hypothesized that each voxel, sampled from a 3 × 3 × 3 mm region of human visual cortex, may have a weak but true bias in its neural or hemodynamic response to different orientations. Such biases could arise from random variations in the distribution or response gain of orientation columns within each voxel. Orientation columns in the monkey typically reveal such spatial variability, and these variations seem to be stable over time[14]. Even if one were to assume that the spatial distribution of orientation columns is perfectly uniform, variability in the distribution of vasculature would lead to uneven hemodynamic sampling across orientation columns, resulting in local biases in orientation preference. We predicted that by pooling together the information from many weakly tuned voxels, the ensemble activity pattern of many voxels may show sharp and stable selectivity for orientation ('ensemble orientation selectivity'). Our experiments showed that different stimulus orientations give rise to distinct patterns of fMRI activity in early human visual areas, which can be accurately decoded by linear pattern analysis techniques. Furthermore, these orientation-selective activity patterns allow for reliable neural decoding of the subjective contents of perception.

[1] ATR Computational Neuroscience Laboratories, 2-2-2 Hikaridai, Keihanna Science City, Kyoto 619-0288, Japan. [2] Psychology Department, Princeton University, Green Hall, Princeton, New Jersey, 08544, USA. [3] Present address: Psychology Department, Vanderbilt University, 301 Wilson Hall, 111 21st Avenue South, Nashville, Tennessee 37203, USA. Correspondence should be addressed to Y.K. (kmtn@atr.jp).

Published online 24 April 2005; doi:10.1038/nn1444

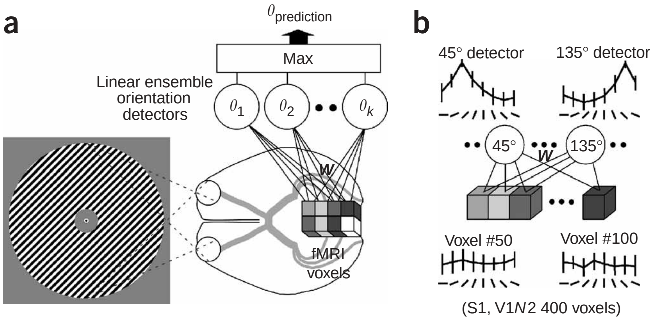

Figure 1 Orientation decoder and ensemble orientation selectivity. (**a**) The orientation decoder predicts stimulus orientation on the basis of fMRI activity patterns. The cubes depict an input fMRI activity pattern obtained while the subject viewed gratings of a given orientation (left). The circles are 'linear ensemble orientation detectors,' each of which linearly combines the fMRI voxel inputs (weighted sum plus bias; bias component not shown). The weights (***W***) are determined by a statistical learning algorithm (linear support vector machine) applied to a training data set, such that the output of each detector becomes largest for its 'preferred orientation' (θ_i). The final unit (rectangle with Max) decides the prediction to be the preferred orientation of the detector with the highest value. (**b**) Orientation selectivity of individual voxels and linear ensemble orientation detectors. The decoder was trained using actual fMRI responses to eight orientations (S1, 400 voxels from V1/V2). Average responses are plotted as a function of orientation for two representative voxels, and for 45° and 135° detectors (error bar, s.d.).

RESULTS

Subjects viewed one of eight possible stimulus orientations while activity was monitored in early visual areas (V1–V4 and MT+) using standard fMRI procedures (3T MRI scanner, spatial resolution 3 × 3 × 3 mm; Methods). For each 16-s 'trial' or stimulus block, a square-wave annular grating (**Fig. 1a**) was presented at the specified orientation (0, 22.5, …, 157.5°), and flashed on and off every 250 ms with a randomized spatial phase to ensure that there was no mutual information between orientation and local pixel intensity. Subjects maintained steady fixation throughout each fMRI run, during which each of the eight stimulus orientations was presented in randomized order. Subjects viewed a total of 20–24 trials for each orientation.

Orientation decoder and ensemble feature selectivity

We constructed an 'orientation decoder' to classify ensemble fMRI activity on individual trials according to stimulus orientation, based on the orientation-selective activity pattern in visual cortex that was obtained from a training data set (**Fig. 1a**). The input consisted of the average response amplitude of each fMRI voxel in the visual area(s) of interest, for each 16-s stimulus trial. In the next layer, 'linear ensemble orientation detectors' for each of the eight orientations received weighted inputs from each voxel and calculated the linearly weighted sum as output. Individual voxel weights for each orientation detector were determined by using a statistical learning algorithm applied to an independent training data set[15]. Voxel weights were optimized so that each detector's output became larger for its preferred orientation than for other orientations (Methods). The final output prediction was made by selecting the most active linear ensemble orientation detector as representing the orientation most likely to be present.

We first trained the orientation decoder using 400 voxels from areas V1/V2 for individual subjects (**Fig. 1b**). Individual voxels showed poor response selectivity for different orientations (**Supplementary Figs. 1** and **2** online). Nonetheless, the output of the linear ensemble orientation detectors, which reflect the weighted sum of many individual voxel responses, had well-tuned responses centered around the preferred orientation of each detector. Furthermore, the detectors showed a graded response that increased according to the similarity of stimulus orientation to their preferred orientation. Because the similarity among orientations was not explicitly specified in the learning procedure, this graded response indicates that similar orientations give rise to more similar patterns of fMRI activity. These results suggest that the ensemble pattern of fMRI activity contains orientation information that greatly exceeds the selectivity of individual voxels.

Orientation decoding accuracy across visual areas

We evaluated if fMRI activity patterns in the human visual cortex are sufficiently reliable to predict what stimulus orientation the subject is viewing on individual trials. We used a cross-validation analysis in which the orientation of each fMRI sample was predicted after the orientation decoder was trained with the remaining samples. Therefore, independent samples were used for training and test. Ensemble fMRI activity in areas V1/V2 led to precise decoding of which of the eight orientations the subject saw on individual stimulus trials (**Fig. 2a**). Decoded orientation responses peaked sharply at the true orientation, with errors, which were infrequent, occurring primarily at neighboring orientations and rarely at orthogonal orientations. The accuracy of these decoded orientation responses was quantified for all four subjects, by calculating the root mean squared error (RMSE) between the true and the predicted orientations, which were 17.9°, 21.0°, 22.2° and 31.2° for subjects S1–S4, respectively (**Fig. 2b**).

In general, orientation decoding performance progressively improved with increasing numbers of voxels, as long as voxels were selected from the retinotopic region corresponding to the stimulated

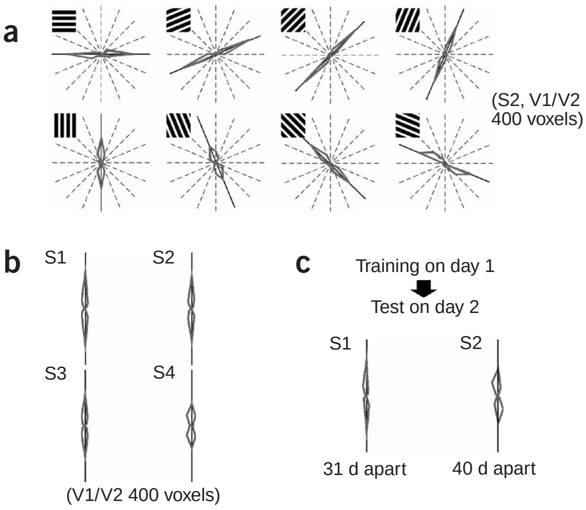

Figure 2 Decoding stimulus orientation from ensemble fMRI activity in the visual cortex. (**a**) Decoded orientation responses for eight orientations. Polar plots indicate the distribution of predicted orientations for each of eight orientations (S2, 400 voxels from V1/V2, 22 samples per orientation). The same values are plotted at symmetrical directions as stimulus orientation repeats every 180°. Solid black lines show the true stimulus orientations. (**b**) Decoded orientation responses for all four subjects (400 voxels from V1/V2, total 160–192 samples for each subject). Results for individual orientations are pooled relative to the correct orientations, and aligned to the vertical line. (**c**) Across-session generalization. Decoded orientation responses were obtained by training a decoder with day 1's data and testing with day 2's data (31 d and 40 d apart for S1 and S2, respectively).

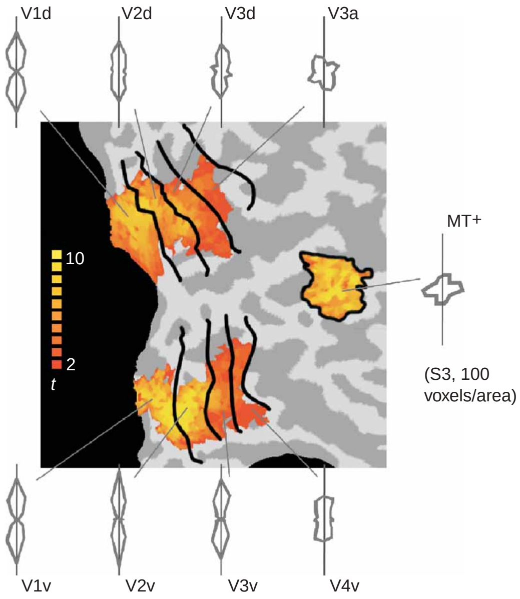

Figure 3 Orientation selectivity across the human visual pathway. Decoded orientation responses are shown for individual visual areas from V1 through V4 and MT+ (S3, 100 voxels per area). The color map indicates *t*-values associated with the responses to the visual field localizer for V1 through V4, and to the MT+ localizer for MT+ (see Methods). The voxels from both hemispheres were combined to obtain the results, though only the right hemisphere is shown. All other subjects showed similar results of progressively diminishing orientation selectivity in higher areas.

visual field. Voxels that showed stronger orientation preference were confined well within retinotopic boundaries of the annular stimulus (**Supplementary Fig. 3**), suggesting the retinotopic specificity of these orientation-selective signals. Consistent with this notion, fMRI activity patterns from unstimulated regions around the foveal representation in V1 and V2 led to chance levels of orientation decoding performance. Although our subjects were well trained at maintaining stable fixation, it is conceivable that different orientations might elicit small but systematic eye movements that could alter the global pattern of cortical visual activity. However, additional control experiments showed that when independent gratings (45° or 135°) were presented simultaneously to each hemifield, visual activity in each hemisphere could accurately decode the orientation of the contralateral stimulus but not the ipsilateral stimulus (96.1% versus 54.9% correct using 200 voxels from V1/V2 for each hemisphere; chance, 50%; **Supplementary Fig. 4**). The independence of orientation information in the two hemispheres cannot be explained in terms of eye movements or any other factors that would lead to global effects on cortical activity.

We further investigated the physiological reliability of these orientation signals in the human visual cortex by testing generalization across separate sessions in two subjects. This was done by training the orientation decoder with fMRI activity patterns from one day and using it to predict perceived orientation with the fMRI data from another day (**Fig. 2c**). The RMSEs for the across-session generalization were 18.9° (31 d apart) and 21.7° (40 d apart) for subjects S1 and S2, respectively, almost as small as those for within-session generalization (17.9° and 21.0°, respectively). The results indicate that these orientation-selective activity patterns reflect physiologically stable response properties across the visual cortex.

The ability to extract robust orientation information from ensemble fMRI activity allowed us to compare orientation selectivity across different human visual areas. Orientation selectivity was most pronounced in early areas V1 and V2, and declined in progressively higher visual areas (**Fig. 3**). All four subjects showed this same trend of diminishing orientation selectivity across retinotopic visual areas (RMSEs, mean ± s.d., for ventral V1, V2, V3 and V4 were 31 ± 4°, 29 ± 4°, 40 ± 8° and 46 ± 4°, respectively). This pattern of orientation selectivity is consistent with monkey data showing poorer orientation selectivity and weaker columnar organization in higher visual areas[10], but has never been shown in the human visual cortex. Unlike areas V1 through V4, human area MT+ showed no evidence of orientation selectivity (53.2 ± 3.7°; chance level, 52.8°), consistent with the idea that this region is more sensitive to motion than to stimulus form.

Source of orientation information

Additional analyses confirmed that this orientation information obtained from the human visual cortex reflects actual orientation-dependent responses. Our linear ensemble orientation detectors were unable to discriminate the orientation of the phase-randomized stimulus gratings on the basis of pixel intensity values alone because orientation is a higher-order property that cannot be expressed by a linearly weighted sum of inputs[16]. However, the decoder composed of linear detectors could classify these images when they received input from intervening orientation filters with nonlinearity, analogous to V1 neurons (**Fig. 4**). The results indicate that ensemble orientation selectivity does not arise from the retinotopic projection of bitmap grating images on the cortex but rather from the orientation information inherent in individual voxels, which can then be pooled together.

What is the pattern of orientation preferences among these voxels that is responsible for such precise ensemble selectivity? We plotted the orientation preference of individual voxels on flattened cortical representations, coloring voxels according to the orientation detector for which each voxel provided the largest weight (**Fig. 5**). Voxel orientation

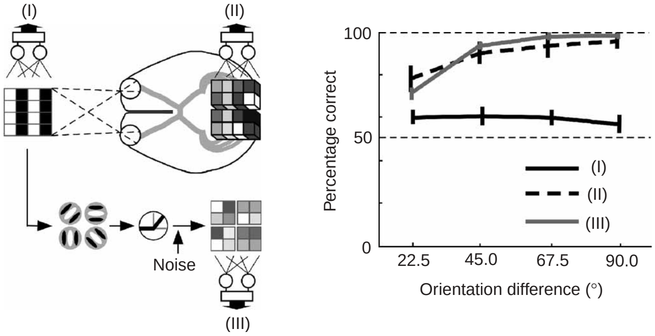

Figure 4 Pairwise decoding performance as a function of orientation difference (all pairs from eight orientations), for grating images (pixel intensities), fMRI images (voxel intensities) and transformed grating images. The gratings (I) were 20 × 20 pixel black-white images with 2–3 stripes. The fMRI images (II) were those obtained in the present study (responses to gratings of eight orientations; 400 voxels from V1/V2). The transformed images (III) were created by linear orientation filtering (Gabor-like filters for four orientations) of the grating images followed by thresholding (nonlinearity) and addition of noise. The orientations of these images were decoded for each pair of orientations (chance level, 50%). For (I) and (III), the average performance with five sets of phase-randomized images is plotted (error bar, s.d.). For (II), the average performance of four subjects is shown. The grating images (I) resulted in poor performance regardless of orientation difference. In contrast, the fMRI images (II) and the transformed grating images (III) both showed performance that improved with orientation difference, reaching near perfect levels at 90°.

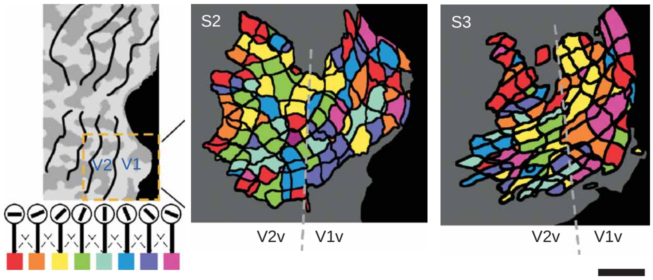

Figure 5 Orientation preference map on flattened cortical surface. The color maps depict the orientation preference of individual voxels on the flattened surface of left ventral V1 and V2 for subjects S2 and S3 (scale bar, 1 cm). Each cell delineated by thick lines is the cross-section of a single voxel (3 × 3 × 3 mm) at the gray-white matter boundary. Voxel colors depict the orientation detector for which each voxel provides the largest weight. The overall color map indicates a template pattern that activates each detector most effectively. The weights were calculated using 400 voxels from V1/V2, including all the quadrants. Other subjects also showed scattered but different patterns of orientation preference. Note that the color map indicates only weak preference for one orientation over others. Simple averaging of the voxels with the same orientation preference led to weak orientation tuning (**Supplementary Fig. 2**), unlike the well-tuned responses of the optimally weighted linear orientation detectors (**Fig. 1b**).

preferences revealed a scattered pattern that was variable and idiosyncratic across subjects. Although some local clustering of the same orientation preference was observed, much of this may reflect spatial blurring resulting from subtle head motion, data reinterpolation required for the Talaraich transformation and the blurred nature of BOLD hemodynamic responses[12,13]. There were no significant differences in the proportion of preferred orientations across different visual areas or different visual field quadrants. Overall, orientation preference maps revealed considerable local variability, indicating that global bias effects due to eye movements or other factors cannot account for the high degree of orientation-selective information that resides in these activity patterns.

Additional analyses showed that orientation selectivity remained equally robust even when the fMRI data underwent normalization to remove differences in mean activity levels across individual activity patterns. Also, differences in mean activity level were small (**Supplementary Fig. 1**), even between canonical and oblique orientations (the oblique effect[17]). Therefore, gross differences in response amplitudes were not a critical source of orientation information for our decoding analysis. We also tested if a global bias for radial orientations might account for our results, as some studies have reported evidence of a bias for orientations radiating outward from the fovea in retinal ganglion cells and V1 neurons of the monkey[18,19]. We removed the global response modulation along the radial dimension from each activity pattern, by dividing voxels into the 16 polar-angle sections obtained from retinotopic mapping, and then normalizing the mean response within each set of iso-polar voxels to the same value for every stimulus trial. After this normalization procedure, orientation selectivity diminished only slightly. The mean RMSEs of four subjects for original and normalized data were 22.7 ± 6.2° and 26.7 ± 7.1°, respectively (200 voxels from V1/V2). Although global factors, such as radial bias, might account for a small degree of the extracted orientation information, local variations in orientation preference seem to provide the majority of the orientation content in these fMRI activity patterns.

The scattered distribution and local variability of orientation preferences across cortex (**Fig. 5**) are consistent with the idea that random variations in the distribution or response strength of individual orientation columns can lead to small local orientation biases that remain detectable at voxel-scale resolutions. To evaluate the viability of this hypothesis, we performed simulations using one-dimensional arrays of orientation columns, which were sampled by coarse-scale voxels and analyzed by our orientation decoder. The array of voxels was allowed to jitter randomly from trial to trial to mimic the effects of small amounts of brain motion. We compared two types of column arrays, one with regularly shifting preferred orientations (**Fig. 6a**) and the other with small random variations in the shifted orientation (**Fig. 6b**), as can be observed in columnar structures in animals. Whereas the regular array showed poor orientation decoding performance, the array with random variation resulted in very similar performance to what was found from actual fMRI activity patterns in the human visual cortex (**Fig. 2**). These results support the hypothesis that a small amount of random variability in the spatial distribution of orientation columns could lead to small local biases in individual voxels that allow for robust decoding of ensemble orientation selectivity.

Mind-reading of attended orientation

Finally, we asked if the ability to characterize brain states corresponding to different stimulus orientations can be extended to the problem of mind-reading, that is, determining a subject's mental state given

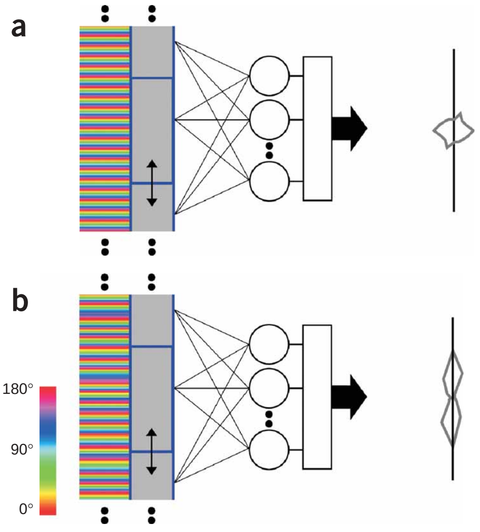

Figure 6 Simulation of one-dimensional array of columns and voxels. (**a**,**b**) Each column was assumed to respond to orientation input according to a Gaussian-tuning function peaking at its preferred orientation (s.d., 45°; noise was added to the output). The preferred orientation shifted by a constant degree (**a**) or by a constant degree plus noise (**b**). In each trial, a single orientation was given as input, and the outputs of 100,000 columns (color band) were sampled by 100 voxels (gray boxes). The actual location of voxel sampling was randomly jittered on each trial (Gaussian distribution with s.d. of a quarter voxel size) to take into account residual head motion. The number of stimulation trials was chosen to match the fMRI experiment. The sampled voxel data were analyzed using the same decoding procedure. As can be seen in the polar plots on the right, orientation can be readily decoded from the irregular array of columns (**b**), but not from the regular array (**a**). Similar results were obtained with a wide range of simulation parameters. Note that if voxel sampling is no longer jittered to mimic minor brain motion, orientation can be decoded even from the regular column array. This is because the high spatial frequency component can still persist after the sampling by large voxels. However, given that pulsatile brain motion and minor head motion cannot be fully eliminated or corrected with 3D alignment procedures, it seems unlikely that such high-frequency information contributes much to the orientation content in our fMRI data.

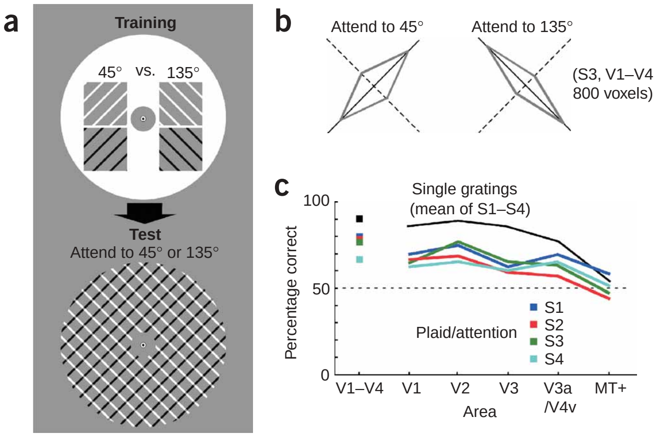

Figure 7 Mind-reading of attended orientation. (**a**) Procedure. First, a decoder was trained using fMRI activity evoked by single gratings to discriminate 45° versus 135°. Black and white gratings (equal contrast) were counterbalanced across trials. In the attention experiment, a plaid pattern composed of two gratings (black-white orientation assignment counterbalanced) was presented. The color of the central fixation spot was changed to indicate to which orientation (45° or 135°) the subject should attend in each trial. The fMRI activity patterns obtained in the attention experiment were classified by the decoder trained with single gratings. (**b**) Performance. Gray lines plot decoded orientation responses for the 'attend to 45°' and 'attend to 135°' conditions (S3, 800 voxels from V1–V4). Solid black lines indicate attended orientations. (**c**) Performance across the human visual pathway. The percentage of correct decoding is plotted by visual area (chance level, 50%; 800 voxels for V1–V4 combined, 200 voxels for V1, V2, V3 and V3a/V4v, and 100 voxels for MT+). Colored lines show the performance of four individual subjects. Black points and lines depict the mean cross-validation performance obtained with single gratings (training session).

knowledge of his or her brain state. Specifically, we tested if the activity patterns evoked by unambiguous single orientations can be used to decode which of two competing orientations is dominant in a person's mind under conditions of perceptual ambiguity. We hypothesized that activity in early human visual areas, which was found here to represent unambiguous stimulus orientations, may subserve a common role in representing the subjective experience of orientation when two competing orientations are viewed. If so, then when subjects are instructed to attend to one of two overlapping gratings (that is, plaid), activity patterns in early visual areas should be biased toward the attended grating.

First, subjects viewed single gratings of 45° or 135° orientation (**Fig. 7a**, top; black and white counterbalanced), and the resulting fMRI activity patterns were used to train the orientation decoder. Then in separate test runs, subjects viewed a plaid stimulus consisting of both overlapping gratings (**Fig. 7a**, bottom). In each 16-s trial, subjects were required to attend to one oriented grating or the other by monitoring for small changes in the width of the bars in the attended grating while ignoring changes in the unattended grating. The fMRI data obtained while subjects viewed the two competing orientations were analyzed using the decoder trained on fMRI activity patterns evoked by the single gratings.

Orientation signals in early human visual areas were strongly biased toward the attended orientation during viewing of the ambiguous stimulus (**Fig. 7b,c**). Even though the same overlapping gratings were presented in the two attentional conditions, fMRI activity patterns in the visual cortex reliably predicted to which of the two orientations the subject was attending on a trial-by-trial basis at overall accuracy levels approaching 80% ($P < 0.0005$ for 4/4 subjects using 800 voxels from V1–V4, chi square test). Analyses of individual visual areas showed that ensemble activity was significantly biased toward the attended orientation in all four subjects for area V1 and also V2 ($P < 0.05$), and in 3/4 subjects for area V3 and areas V3a/V4v combined.

We did an additional control experiment to address whether eye movements, orthogonal to the attended orientation, might account for the enhanced responses to the attended grating by inducing retinal motion. The visual display was split into left and right halves, and activity from corresponding regions of the contralateral visual cortex was used to decode the attended orientation in each visual field (**Supplementary Fig. 4**). Even when the subject was instructed to pay attention to different orientations in the plaids of the left and right visual fields simultaneously, cortical activity led to accurate decoding of both attended orientations. Because eye movements would bias only one orientation in the whole visual field, these results indicate that the attentional bias effects in early visual areas are not due to retinal motion induced by eye movements.

The robust effects found in V1 and V2 suggest that top-down voluntary attention acts very early in the processing stream to bias orientation-selective signals. Although previous studies have reported evidence of feature-based attentional modulation in the visual cortex[20–25], our results provide novel evidence that top-down attention can bias orientation signals at the earliest stage of cortical processing when two competing stimuli are entirely overlapping. These results suggest that feedback signals to V1 and V2 may have an important role in voluntary feature-based attentional selection of orientation signals.

DISCUSSION

We have shown that fMRI activity patterns in the human visual cortex contain reliable orientation information that allows for detailed prediction of perceptual and mental states. By combining an ensemble of weakly orientation-selective fMRI voxels, we could accurately differentiate subtle variations in perceived stimulus orientation on a trial-by-trial basis. Moreover, activity patterns evoked by unambiguous stimuli could reliably predict which of two competing orientations was the focus of a subject's attention. These results demonstrate a tight coupling between brain states and subjective mental states.

Models of human visual perception commonly assume orientation-selective units, similar to those found in animals, to account for a variety of psychophysical data[26]. However, neurophysiological evidence to support the existence of such mechanisms in the human brain has been indirect[27–29]. Here, we found that early human visual areas were indeed highly orientation selective, and that ensemble orientation selectivity was most pronounced in areas V1 and V2 and progressively weaker in higher areas, consistent with neurophysiological data in monkeys[10]. Our results suggest that orientation selectivity in the human visual system may closely resemble that of nonhuman primates.

Previous neuroimaging studies have used multi-voxel pattern analysis to reveal broadly distributed object representations extending over 1–5 cm regions of the ventral temporal cortex[7]. Our approach for measuring ensemble feature selectivity suggests that multi-voxel analyses may also be effective at extracting feature-tuned information at much finer scales of cortical representation, by pooling together weak feature-selective signals in each voxel, which may arise from variability in the distribution of cortical feature columns or their vascular supply. If this is indeed the case, then the approach proposed here could be used to investigate a variety of feature domains, such as color selectivity, motion selectivity or feature tuning in other sensory and motor modalities, and thus may provide a useful bridge between animal studies and human neuroimaging studies.

We emphasize the importance of using linear approaches to study ensemble feature selectivity. Our analysis relied on a linear weighting procedure to measure the orientation-selective information inherent in each voxel and to pool together the ensemble information from many voxels. In contrast, nonlinear pattern analysis techniques could extract orientation information even if every individual voxel lacked any orientation selectivity. Flexible nonlinear approaches would allow for nonlinear interactions between input voxels; these could be used to construct nonlinear oriented filters to decode orientation on the basis of pixel intensity information alone. Therefore, nonlinear methods may spuriously reflect the feature-tuning properties of the pattern analysis algorithm rather than the tuning properties of individual units within the brain. For these reasons, it is important to restrict the flexibility of pattern analysis methods when measuring ensemble feature selectivity.

Finally, our method of measuring ensemble feature selectivity proved highly effective at decoding mental states from measured brain states. Our results on the mind-reading of attended orientation address an ongoing debate regarding the role of early visual areas and primary visual cortex in visual awareness[30,31]. Voluntary top-down attention strongly biased the pattern of orientation responses at the earliest stages of cortical processing in V1 and V2. These results suggest that feedback projections to these early visual areas may be important in feature-based attentional selection and in maintaining a particular orientation in awareness. More generally, the mind-reading approach presented here provides a potential framework for extending the study of the neural correlates of subjective experience[32]. True understanding of the neural basis of subjective experience should allow for reliable prediction of a person's mental state based solely on measurements of his or her brain state. By analyzing ensemble activity from human cortex, we were able to extract information about the person's subjective mental state under conditions of ambiguity. Our approach may be extended to studying the neural basis of many types of mental content, including a person's awareness, attentional focus, memory, motor intention and volitional choice. Further development of such mind-reading approaches may eventually lead to potential applications for non-invasive brain-machine interface[33,34] by providing effective procedures to translate brain activity into mental contents.

METHODS

Subjects. Four healthy adults with normal or corrected-to-normal visual acuity participated in this study. All subjects gave written informed consent. The study was approved by the Institutional Review Panel for Human Subjects at Princeton University.

Experimental design and stimuli. Visual stimuli were rear-projected onto a screen in the scanner bore using a luminance-calibrated LCD projector driven by a Macintosh G3 computer.

Each experimental run to measure fMRI responses to gratings consisted of a series of 16-s stimulus trials (with no intervening rest periods), plus 16-s fixation-rest periods at the beginning and at the end of each series.

In the eight-orientation experiment, each run had eight trials for eight different orientations (0°, 22.5°…157.5°). In each trial, a square-wave annular grating of a given orientation (~100% contrast, 1.5 cycles per degree; 1.5°–10° of eccentricity) was flashed at 2 Hz (on/off for 250 ms). The spatial phase of the grating was randomized in each frame so that the pixel intensity did not carry information about orientation. Because of the long temporal interval between flashed gratings, apparent motion among the gratings was hardly visible[35]. The subject passively viewed the stimulus while maintaining fixation on a central fixation spot. The order of orientation conditions was randomized in each run. Each subject performed 20–24 runs for a total of 20–24 trials per orientation.

In the attention experiment, we interleaved two types of runs, training and test runs. In each trial of the training runs, either a 45° or 135° grating (black or white stripes; 0.75 cycles per degree) was flashed on a gray background at 0.5 Hz (on for 1,750 ms, off for 250 ms) with a randomized spatial phase. The subject monitored the thickness of the stripes, and reported whether they were thick or thin (one- or two-pixel difference) by pressing a key within each 2-s stimulus frame. A single run had 16 trials, and 6 runs were repeated in each subject. The order of orientation (45° or 135°) and stripe color (black or white) was randomized.

In the test runs of the attention experiment, two overlapping gratings of 45° and 135° (white and black, or black and white; counterbalanced across trials) were flashed with the same time course as in the training runs. The intersections of the white and black stripes were gray to enhance the perceptual segregation of the two gratings. During each 16-s trial, the color of the fixation spot was either red or green (randomized), indicating which grating (45° or 135°) the subject should monitor for changes in thickness. (The color-orientation rule was reversed in the middle of the experiment.) A single run had 16 trials, and 6 runs were repeated in each subject. The overall performance of the task was 79 ± 5% correct with a 50% chance level. Thus, the task was difficult enough to restrict attention to one of the gratings.

In the same session, subjects viewed a reference stimulus to localize the retinotopic regions corresponding to the stimulated visual field. The 'visual field localizer' composed of high-contrast dynamic random dots was presented in an annular region for 12-s periods, interleaved with 12-s rest/fixation periods, while the subject maintained fixation. We used a smaller annular region for the visual field localizer (2°–9° of eccentricity) than for the gratings (1.5°–10°) to avoid selecting voxels corresponding to the stimulus edges, which may contain information irrelevant to grating orientation. In separate sessions, standard retinotopic mapping[12,36] and MT+ localization procedures[37–39] were done to delineate visual areas on flattened cortical representations.

MRI acquisition. Scanning was performed on a 3.0-Tesla Siemens MAGNETOM Allegra scanner using a standard head coil at the Center for the Study of Brain, Mind and Behavior, Princeton University. A high-resolution T1-weighted anatomical scan was acquired for each participant (FOV 256 × 256, 1 mm^3 resolution). To measure BOLD contrast, standard gradient-echo echoplanar imaging parameters were used to acquire 25 slices perpendicular to the calcarine sulcus to cover the entire occipital lobe (TR, 2,000 ms; TE, 30 ms; flip angle, 90°; slice thickness, 3 mm; in-plane resolution, 3 × 3 mm). A custom-made bite bar was used to minimize head motion.

Functional MRI data preprocessing. All fMRI data underwent three-dimensional (3D) motion correction using automated image registration software[40], followed by linear trend removal. No spatial or temporal smoothing was applied. The fMRI data were aligned to retinotopic mapping data collected in a separate session, using Brain Voyager software (Brain Innovation). Automated alignment procedures were followed by careful visual inspection and manual fine-tuning at each stage of alignment to correct for misalignment error. Rigid-body transformations were done to align fMRI data to the within-session 3D anatomical scan, and next to align these data to retinotopy data. After our alignment procedure, any residual misalignment between fMRI scans collected across the two sessions appeared very small (less than 1 mm) and were likely of comparable order of magnitude to brain motion resulting from respiration, heart rate and residual head movement while subjects were stabilized with a bite bar. After across-session alignment, fMRI data underwent Talairach transformation and reinterpolation using 3 × 3 × 3 mm voxels. This transformation allowed us to restrict voxels around the gray-white matter boundary and to delineate individual visual areas on flattened cortical representations. However, these procedures involving motion correction and interpolation of the raw fMRI data may have resulted in the reduction of orientation information that may be contained in fine-scale activity patterns.

Voxels used for orientation decoding analysis were selected on the cortical surface of V1 through V4 and MT+. First, voxels near the gray-white matter boundary were identified within each visual area using retinotopic maps delineated on a flattened cortical surface representation. Then, the voxels were sorted according to the responses to the visual field localizer (V1–V4) or to the MT+ localizer. We used 200 voxels for each of areas V1–V4 (100 voxels when dorsal and ventral parts were separately analyzed) and 100 voxels for MT+ by selecting the most activated voxels.

The data samples used for orientation decoding analysis were created by shifting the fMRI time series by 4 s to account for the hemodynamic delay, and then averaging the MRI signal intensity of each voxel for each 16-s stimulus trial. Response amplitudes of individual voxels were normalized relative to the average of the entire time course within each run (excluding the rest periods at the beginning and the end) to minimize baseline differences across runs. The resulting activity patterns were labeled according to their corresponding stimulus orientation and served as input to the orientation decoder analysis.

Decoding analysis. The calculation performed by each linear ensemble orientation detector with preferred orientation θ_k can be expressed by a linear function of voxel inputs $\mathbf{x} = (x_1, x_2, \ldots, x_d)$ ('linear detector function')

$$g_{\theta_k}(\mathbf{x}) = \sum_{i=1}^{d} w_i x_i + w_0$$

where w_i is the weight of voxel i and w_0 is the bias. To achieve this function for each orientation, we first calculated linear discriminant functions for pairs of orientations using machine learning–pattern recognition algorithms. Linear support vector machines[15] (SVM) were used to obtain the results presented here, although other algorithms, such as Fisher's linear discriminant method (combined with principal component analysis) and Perceptrons, could be used to yield similar results. The discriminant function, $g_{\theta_k\theta_l}(\mathbf{x})$ for the discrimination of orientations θ_k and θ_l , is expressed by a weighted sum of voxel inputs plus bias, and satisfies

$$g_{\theta_k\theta_l}(\mathbf{x}) > 0 \ (\mathbf{x} \text{ is fMRI activity induced by orientation } \theta_k)$$
$$g_{\theta_k\theta_l}(\mathbf{x}) < 0 \ (\mathbf{x} \text{ is fMRI activity induced by orientation } \theta_l).$$

Using a training data set, a linear SVM finds optimal weights and bias for the discriminant function. After the normalization of the weight vectors, the pairwise discriminant functions comparing θ_k and the other orientations were simply added to yield the linear detector function

$$g_{\theta_k}(\mathbf{x}) = \sum_{m \neq k} g_{\theta_k\theta_m}(\mathbf{x}).$$

This linear detector function becomes larger than zero when the input $\mathbf{x}$ (in the training data set) is an fMRI activity pattern induced by orientation θ_k. In our method, orientation is treated as a categorical variable, and no similarity among orientations is assumed.

To evaluate orientation decoding performance, we performed a version of cross-validation by testing the fMRI samples in one run using a decoder trained with the samples from all other runs. This training-test set was repeated for all runs ('leave one run out' cross-validation). We used this procedure to avoid using the samples in the same run both for training and test, as they are not independent because of the normalization of voxel intensity within each run.

Note: Supplementary information is available on the Nature Neuroscience website.

ACKNOWLEDGMENTS

We thank D. Remus and J. Kerlin for technical support, the Princeton Center for Brain, Mind and Behavior for MRI support and J. Haxby, D. Heeger and A. Seiffert for comments. This work was supported by grants from Japan Society for the Promotion of Science and National Institute of Information and Communications Technology to Y.K., and grants R01-EY14202 and P50-MH62196 from the US National Institutes of Health to F.T.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 3 January; accepted 29 March 2005
Published online at http://www.nature.com/natureneuroscience/

1. Hubel, D.H. & Wiesel, T.N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol. (Lond.)* **160**, 106–154 (1962).
2. Hubel, D.H. & Wiesel, T.N. Receptive fields and functional architecture of monkey striate cortex. *J. Physiol. (Lond.)* **195**, 215–243 (1968).
3. Blasdel, G.G. Orientation selectivity, preference, and continuity in monkey striate cortex. *J. Neurosci.* **12**, 3139–3161 (1992).
4. Bartfeld, E. & Grinvald, A. Relationships between orientation-preference pinwheels, cytochrome oxidase blobs, and ocular-dominance columns in primate striate cortex. *Proc. Natl. Acad. Sci. USA* **89**, 11905–11909 (1992).
5. Paradiso, M.A. A theory for the use of visual orientation information which exploits the columnar structure of striate cortex. *Biol. Cybern.* **58**, 35–49 (1988).
6. Pouget, A., Dayan, P. & Zemel, R. Information processing with population codes. *Nat. Rev. Neurosci.* **1**, 125–132 (2000).
7. Haxby, J.V. *et al.* Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* **293**, 2425–2430 (2001).
8. Cox, D.D. & Savoy, R.L. Functional magnetic resonance imaging (fMRI) 'brain reading': detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage* **19**, 261–270 (2003).
9. Carlson, T.A., Schrater, P. & He, S. Patterns of activity in the categorical representations of objects. *J. Cogn. Neurosci.* **15**, 704–717 (2003).
10. Vanduffel, W., Tootell, R.B., Schoups, A.A. & Orban, G.A. The organization of orientation selectivity throughout macaque visual cortex. *Cereb. Cortex* **12**, 647–662 (2002).
11. Kim, D.S., Duong, T.Q. & Kim, S.G. High-resolution mapping of iso-orientation columns by fMRI. *Nat. Neurosci.* **3**, 164–169 (2000).
12. Engel, S.A., Glover, G.H. & Wandell, B.A. Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cereb. Cortex* **7**, 181–192 (1997).
13. Malonek, D. & Grinvald, A. Interactions between electrical activity and cortical microcirculation revealed by imaging spectroscopy: implications for functional brain mapping. *Science* **272**, 551–554 (1996).
14. Shtoyerman, E., Arieli, A., Slovin, H., Vanzetta, I. & Grinvald, A. Long-term optical imaging and spectroscopy reveal mechanisms underlying the intrinsic signal and stability of cortical maps in V1 of behaving monkeys. *J. Neurosci.* **20**, 8111–8121 (2000).
15. Vapnik, V.N. *Statistical Learning Theory* (Wiley, New York, 1998).
16. Minsky, L.M. & Papert, S.A. *Perceptrons – Expanded Edition: An Introduction to Computational Geometry* (MIT Press, Boston, 1987).
17. Furmanski, C.S. & Engel, S.A. An oblique effect in human primary visual cortex. *Nat. Neurosci.* **3**, 535–536 (2000).
18. Bauer, R. & Dow, B.M. Complementary global maps for orientation coding in upper and lower layers of the monkey's foveal striate cortex. *Exp. Brain Res.* **76**, 503–509 (1989).
19. Schall, J.D., Perry, V.H. & Leventhal, A.G. Retinal ganglion cell dendritic fields in old-world monkeys are oriented radially. *Brain Res.* **368**, 18–23 (1986).
20. Treue, S. & Maunsell, J.H. Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature* **382**, 539–541 (1996).
21. Treue, S. & Martinez Trujillo, J.C. Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* **399**, 575–579 (1999).
22. Treue, S. Visual attention: the where, what, how and why of saliency. *Curr. Opin. Neurobiol.* **13**, 428–432 (2003).
23. Roelfsema, P.R., Lamme, V.A. & Spekreijse, H. Object-based attention in the primary visual cortex of the macaque monkey. *Nature* **395**, 376–381 (1998).
24. Watanabe, T. *et al.* Task-dependent influences of attention on the activation of human primary visual cortex. *Proc. Natl. Acad. Sci. USA* **95**, 11489–11492 (1998).
25. Saenz, M., Buracas, G.T. & Boynton, G.M. Global effects of feature-based attention in human visual cortex. *Nat. Neurosci.* **5**, 631–632 (2002).
26. Wilson, H.R., Levi, D., Maffei, L., Rovamo, J. & DeValois, R. in *Visual Perception: The Neurophysiological Foundations* (eds. Spillman, L. & Werner, J.S.) 231–272 (Academic, San Diego, 1990).
27. Tootell, R.B. *et al.* Functional analysis of primary visual cortex (V1) in humans. *Proc. Natl. Acad. Sci. USA* **95**, 811–817 (1998).
28. Boynton, G.M. & Finney, E.M. Orientation-specific adaptation in human visual cortex. *J. Neurosci.* **23**, 8781–8787 (2003).
29. Kamitani, Y. & Shimojo, S. Manifestation of scotomas created by transcranial magnetic stimulation of human visual cortex. *Nat. Neurosci.* **2**, 767–771 (1999).
30. Rees, G., Kreiman, G. & Koch, C. Neural correlates of consciousness in humans. *Nat. Rev. Neurosci.* **3**, 261–270 (2002).
31. Tong, F. Primary visual cortex and visual awareness. *Nat. Rev. Neurosci.* **4**, 219–229 (2003).
32. Koch, C. *The Quest for Consciousness: A Neurobiological Approach* (Roberts, Englewood, Colorado, USA, 2004).
33. Donoghue, J.P. Connecting cortex to machines: recent advances in brain interfaces. *Nat. Neurosci.* **5** (suppl.): 1085–1088 (2002).
34. Wolpaw, J.R. & McFarland, D.J. Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. *Proc. Natl. Acad. Sci. USA* **101**, 17849–17854 (2004).
35. Kahneman, D. & Wolman, R.E. Stroboscopic motion: Effects of duration and interval. *Percept. Psychophys.* **8**, 161–164 (1970).
36. Sereno, M.I. *et al.* Borders of multiple visual area in humans revealed by functional magnetic resonance imaging. *Science* **268**, 889–893 (1995).
37. Zeki, S. *et al.* A direct demonstration of functional specialization in human visual cortex. *J. Neurosci.* **11**, 641–649 (1991).
38. Watson, J.D. *et al.* Area V5 of the human brain: evidence from a combined study using positron emission tomography and magnetic resonance imaging. *Cereb. Cortex* **3**, 79–94 (1993).
39. Tootell, R.B. *et al.* Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *J. Neurosci.* **15**, 3215–3230 (1995).
40. Woods, R.P., Grafton, S.T., Holmes, C.J., Cherry, S.R. & Mazziotta, J.C. Automated image registration: I. General methods and intrasubject, intramodality validation. *J. Comput. Assist. Tomogr.* **22**, 139–152 (1998).