Early visual brain areas reflect the percept of an ambiguous scene

Lauri Parkkonen[a,1], Jesper Andersson[a,b], Matti Hämäläinen[c,d], and Riitta Hari[a,e,1]

[a]Brain Research Unit, Low Temperature Laboratory, Helsinki University of Technology, FIN-02015 TKK, Finland; [b]Oxford Centre for Functional Magnetic Resonance Imaging of the Brain, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, United Kingdom; and [c]Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA 02129; [d]Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA 02139; and [e]Department of Clinical Neurophysiology, Helsinki University Central Hospital, FIN-00290 Helsinki, Finland

When a visual scene allows multiple interpretations, the percepts may spontaneously alternate despite the stable retinal image and the invariant sensory input transmitted to the brain. To study the brain basis of such multi-stable percepts, we superimposed rapidly changing dynamic noise as regional tags to the Rubin vase-face figure and followed the corresponding tag-related cortical signals with magnetoencephalography. The activity already in the earliest visual cortical areas, the primary visual cortex included, varied with the perceptual states reported by the observers. These percept-related modulations most likely reflect top-down influences that accentuate the neural representation of the perceived object in the early visual cortex and maintain the segregation of objects from the background.

bistable perception | frequency tagging | human | magnetoencephalography | visual system

Like other senses, the visual system receives only sparse information about the surrounding world and resorts to *a priori* assumptions to infer the physical structure underlying a visual scene. This process yields a unique interpretation, i.e., a stable percept, for most visual scenes we encounter. However, when we are confronted with an ambiguous image that is compatible with multiple approximately equally probable interpretations, the percepts may start to spontaneously alternate. This perceptual instability is stochastic, meaning that the durations of stable percepts are unpredictable.

Studies of the neural signatures of perceptual switching (1) have not resolved at which stage in the hierarchy of visual cortical areas the activity correlates with the perceptual state when the subject is viewing an ambiguous figure. We now report such percept-dependent modulations already in the early visual cortices.

Much of the current evidence supporting the involvement of the early visual cortical areas in bistable perception is based on binocular rivalry (2, 3), the spontaneously alternating perceptual dominance of one eye when the eyes receive incongruent images. However, these results are difficult to interpret because they concern competition between the two eyes and the two perceptual contents. Ambiguous figures, presented identically to both eyes, are more closely related to the naturally occurring perceptual ambiguity, but with such stimuli the link between perceptual and neural states has been difficult to establish.

We introduce here a novel method to follow the cortical signals elicited by the well known Rubin bistable vase-face figure to investigate whether neural activity during the alternating percepts would differ already in the early visual areas. To this end, we tagged the figure with subtle dynamic noise that oscillated at 12 Hz in the vase and at 15 Hz in the face regions [see supporting information (SI) Movie S1] and monitored the brain signals at these frequencies with magnetoencephalography (MEG). Compared with the whole stimulus flickering at distinct frequencies used to study binocular rivalry (4–7), the dynamic noise used here distorted the image only minimally, preserving both the percepts and their spontaneous switching. Importantly, the subjects were not able to distinguish between the two noise oscillation rates.

As a result of clearly different recovery times in different brain areas (8), stimulus-locked activity at 12 to 15 Hz propagated synchronously only to the earliest visual cortices, whose activity could thus be probed by measuring the corresponding MEG signals.

Results

Eight of the nine subjects reported spontaneous alternation of the two percepts when they were viewing the noise-tagged face-vase figure during the MEG measurement. As reported previously (1, 9), the durations of percepts followed a gamma distribution, here with a mean of 4.2 s (Fig. 1*A*). After excluding percepts shorter than 2 s and epochs contaminated by eye blinks, 28 to 101 MEG epochs per subject were available for further analysis.

Eye-movement tracking revealed indistinguishable gaze distributions during the two percepts (Fig. 1*B*), thereby confirming the similarity of retinal images during both percepts.

In all subjects, amplitude spectra of occipital MEG sensors exhibited clear peaks (at least twice the mean spectral density between 12.5 and 14.5 Hz) at the tag frequencies; Fig. 2*B* shows the spectrum for one subject (for spectra for all subjects, see Fig. S1).

The averaged instantaneous power of occipital MEG sensors in the subject of Fig. 2*C* shows a clear dominance of the vase tag (12 Hz) signal during the vase percept and of the face tag (15 Hz) signal during the faces percept.

Cortical source estimates indicated generation of the tagged MEG signals mostly in the peri-calcarine cortex (Fig. 3*A*). All eight subjects had statistically significant ($p < 0.0051$) activity at both tag frequencies at the posterior part of the calcarine sulcus, close to the occipital pole, corresponding to small eccentricities of the retinal input. The present data do not allow demarcation between V1 and V2 cortices, both of which were probably activated. In most subjects, the lateral occipital (LO) cortex also displayed tagged signals, although considerably weaker than those from the calcarine area. Because of the low signal amplitudes, the LO activity was not analyzed further.

Group-level analysis of the signal power within our region of interest (ROI; Fig. 3*A*) in the calcarine cortex indicated an enhancement of the vase tag (12 Hz) for the vase percept and an enhancement of the face tag (15 Hz) for the faces percept (Fig.
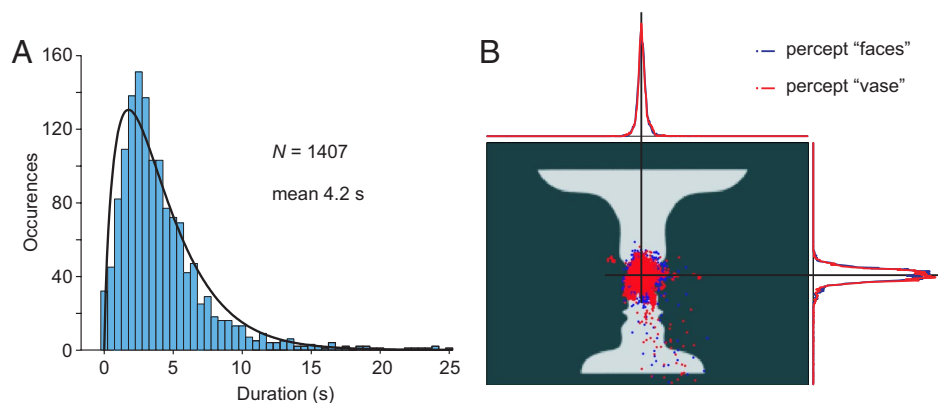
Fig. 1. Behavioral and eye tracking data. (*A*) The durations of the alternating percepts reported by the subjects. The black line shows the fitted gamma distribution. (*B*) Distributions of eye fixations for the two percepts; data from one subject.

3*B*), consistent with the analysis performed on occipital sensor signals (Fig. 2*C*). However, neither tag was abolished during the opposite percept. The average modulation of the tag-driven signals, relative to the mean amplitude of the signal, was approximately 12%.

Importantly, the balance of the tag-related brain signals followed the behaviorally reported percept. Within the ROI, the amplitude ratio of the vase tag to the face tag was larger during the vase than the face percept in all eight subjects ($P = 0.0039$, binomial test; see Fig. 3*C*).

Discussion

Our data demonstrate that the changing percepts of an ambiguous figure are associated with changes in the neural activity in the early visual cortical areas: tagging of the vase and face regions of Rubin ambiguous figure elicited well discernible oscillatory signals over the posterior parts of the brain, and the strengths of these signals varied in accordance with the prevailing percept. Eye movement monitoring confirmed that the variation of tag signals and percepts cannot be explained by changes of the retinal images.

The relatively high frequencies of the noise tags superimposed on the stimulus were selected specifically to probe the early visual areas that are more resilient than the later visual cortices to high stimulation rates (8). Although stimuli flickering at lower

frequencies have been successfully used for recording cortical activity, e.g., in the LO cortex (10), the wide network of visual areas activated by such low-frequency tags would have complicated the interpretation of our data.

The dynamic noise tagging was not perceptually salient. Most importantly, it retained the perceptual switching. The method might thus be further exploited in studies of the early visual areas with time-sensitive methods, such as MEG and EEG.

The tag-related, percept-dependent MEG signals were generated in the early visual areas, predominantly in V1. The observed modulation (by an average of 12%) of the signals is in line with findings that, during binocular rivalry, only a fraction (18%) of monkey V1/V2 neurons follow the percept (3).

As our retinal stimulus was invariant throughout the experiments, the percept-dependent changes we observed are likely top-down modulation of the V1 region from extrastriate visual areas. In monkey V4 and MT, the percentage of neurons following the dominant percept in binocular rivalry is approximately twice as large as that in V1/V2 (3), suggesting that those areas either directly influence the V1/V2 activity via the recurrent connections or share with V1/V2 some, possibly non-visual, information that alters the perceptual bias (11, 12). Such steering of perception could come from brain regions involved in planning and selection of goal-directed behavior (13–15). In addition, pre-stimulus activity fluctuations in the fusiform face area

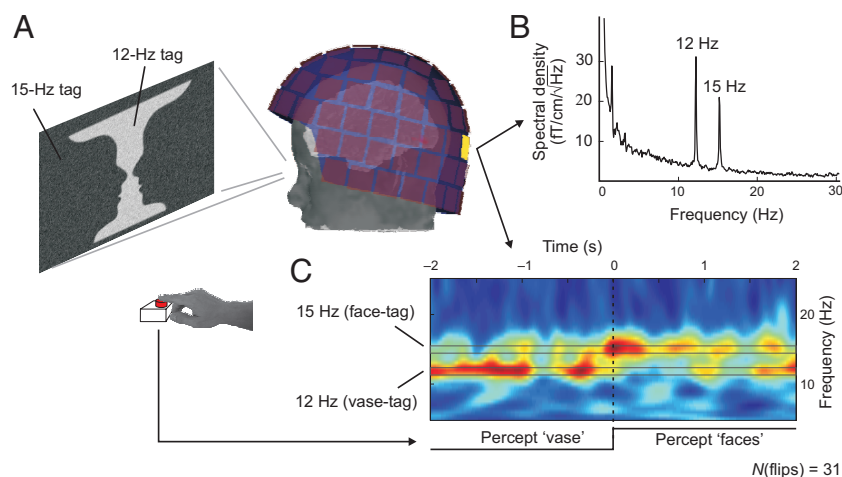
Fig. 2. Experimental setup and MEG signals from one subject. (*A*) Dynamic noise at two update frequencies superimposed on Rubin face-vase figure to tag the face and vase regions. (*B*) Amplitude spectrum computed from the MEG signals in one occipital sensor. (*C*) The time-frequency plot of the tag-related signals averaged across 31 perceptual flips. The warm colors indicate larger power.
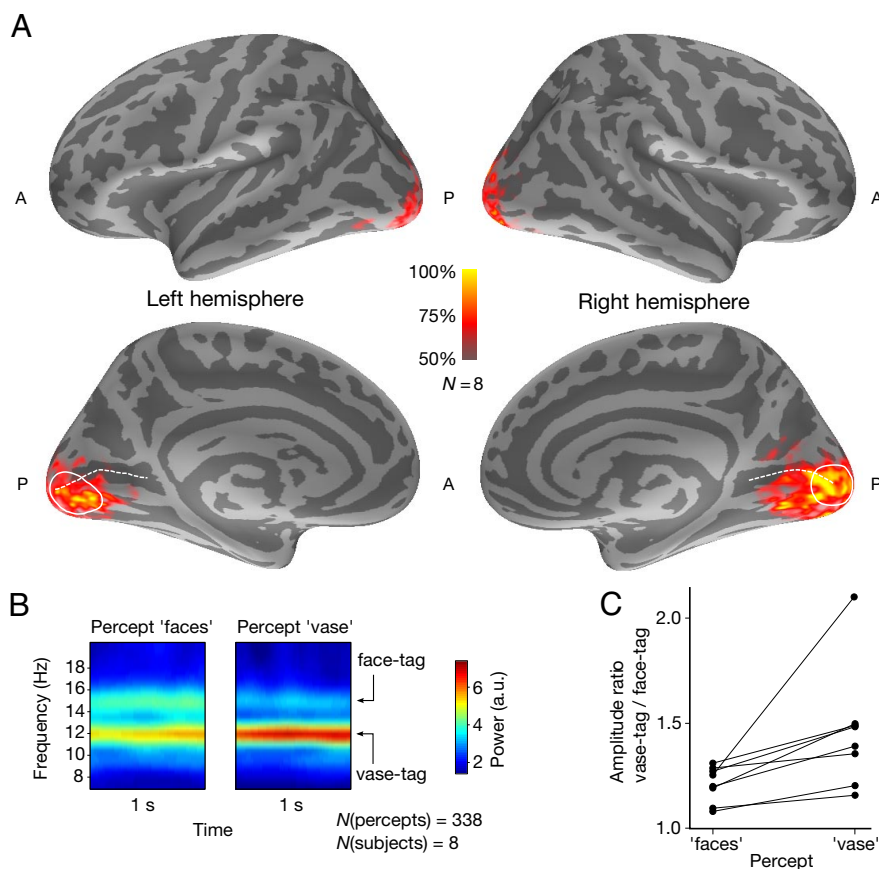
Fig. 3. Cortical activity. (*A*) Tag-related cortical activity detected by MEG. The colors on the inflated average brain indicate the percentage of subjects with statistically significant (*p* < 0.0051) activity at either tag frequency. The dashed white line denotes the calcarine sulcus, and the solid line encloses the ROI. (*B*) The time-frequency representations of the tag-related signals within the ROI during the two percepts; mean activity across subjects. (*C*) Tag amplitude ratios within the ROI for the faces and vase percepts in all eight subjects.

have recently been shown to bias the subsequent perceptual outcome of the presentation of the Rubin vase-face figure (16).

Our result, based on the quantification of the relative changes of the signals associated with the two perceptual states, implies enhanced tag-related activity in the V1/V2 cortices during the corresponding percept. This relative enhancement does not contradict functional magnetic resonance imaging findings of suppressed V1 activity in association with the perceptual switches *per se* (1) or during bistable perceptual grouping (17). Moreover, figure-ground alternation in our stimulus differs qualitatively from bistable perceptual grouping and thus may engage different neural mechanisms.

Activity in the early visual cortices is known to reflect the perceived size of an object, independently of its size on the retina (18), and the dominant percept during binocular rivalry (3, 19, 20), as well as the visual context of the target object (21). Behavioral experience and perceptual saliency also modulate the activity in those areas (22). Our results support the view emerging from these studies that the activity in early visual areas is modulated according to the perceptual state and is not only a simple feed-forward reflection of the retinal input (23).

All sensory input is inherently ambiguous (24): even during seemingly unambiguous percepts, neuronal modulations in the early visual areas likely help to maintain the segregation of figure and ground (25). The spontaneous perceptual switching may manifest a mechanism that prevents the visual system from being trapped into an interpretation that is later invalidated by subtle changes in the stimulus or its context (13).

Materials and Methods

Subjects. Nine healthy human subjects participated in the study after informed consent. One subject did not experience perceptual switching; the data from the other eight subjects (age range, 23–41 y; mean age, 30.6 y; five men) were analyzed. The MEG recordings had prior approval by the local ethics committee.

Stimuli. The dynamic noise for tagging the stimulus figure was created by adding random values from −64 to +64 to the intensity (range, 0–255) of each pixel in the original image (640 × 480 pixels). In test recordings on three subjects with noise-tagged rectangles as stimuli, update rates between 10 and 20 Hz provided the most distinct spectral peaks in MEG. The chosen 12- and 15-Hz rates were realized as a sequence of 120 images in which the random noise pattern changed every fifth frame in the vase region and every fourth frame in the face region (see Movie S1). This sequence was seamlessly looped at 60 frames per second (Presentation software; Neurobehavioral Systems). For the percept-independent average, the MEG system received a trigger pulse from the stimulus presentation system once every 60 frames. The image was projected by a triple-DLP projector (VistaPro; Christie Digital Systems) onto a screen where it subtended a horizontal visual angle of 24°; a single pixel was 0.076° in extent.

MEG Measurements. MEG was recorded in a dimly lit magnetically shielded room with a 306-channel whole-scalp device comprising 204 planar gradiometers and 102 magnetometers (Elekta Neuromag Oy). Four indicator coils attached to the scalp were used to measure the location of the head in the MEG sensor array. The MEG signals and a diagonal electro-oculogram were filtered to 0.1–170 Hz and were synchronously sampled at 600 Hz, together with the finger position signal and stimulus sequence trigger.

The subjects were asked to fixate between the two "noses" of the vase-face figure and to indicate the percept by holding down the right index finger for the entire duration of the face percept; the finger position was detected with
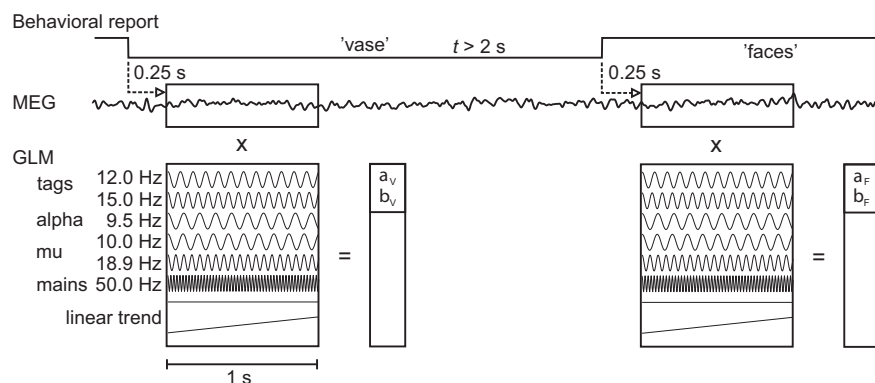
Fig. 4. Estimation of the tag amplitudes (a_F and b_F for the faces percept; a_V and b_V for the vase percept) from the MEG signals. A general linear model was fitted to the cortical estimate of MEG after each reported perceptual flip; the regressors include both tags and subject-specific spontaneous brain rhythms.

a silent optical switch. Before data collection, subjects practiced the task until they felt they could report the change of the percept reliably and swiftly.

MEG was recorded continuously for 10 min, followed by a 1-min recording without stimuli in which the subject continued to fixate on the center of the screen.

Eye Tracking. Eye fixation was verified in one subject by using an eye tracker (SensoMotoric Instruments) in the MEG environment. The gaze direction data, calibrated with nine points, were sampled at 50 Hz. The samples were classified according to the reported percept, and the spatial distributions were estimated.

MEG Data Analysis. External interference on the MEG signals was suppressed with signal-space separation (26). Tag-related MEG signals were identified from amplitude spectra: the data were first low-pass filtered at 50 Hz, then down-sampled to 150 samples/s, and the amplitudes of half-overlapping Hanning-windowed 4,096-point fast-Fourier transforms were averaged across the recording to yield a frequency resolution of 0.073 Hz.

To estimate the cortical areas giving rise to the tag-related MEG signals, the continuous recording was split into 1-s segments that were averaged, irrespective of the reported percept, across the 10-min recording to yield an oscillatory evoked response with a high signal-to-noise ratio. The trigger ensured a consistent phase of the stimulus tag with respect to the averaging window.

For MEG source analysis, the geometry of the inner skull surface (for a boundary-element conductor model) and that of the cortical surface were determined from individual magnetic resonance images by using FreeSurfer software (27, 28). These structural data were used to compute a depth-weighted and noise-normalized minimum-norm estimate (29) on the cortical surface with an approximate 5-mm distance between adjacent points. The currents were then aligned to be approximately perpendicular to the cortical mantle (i.e., loose orientation constraint approach [30]; variance ratio of 3.3:1 between currents perpendicular and tangential to the cortex at each source location).

The amplitudes of the tag-related signals were estimated in temporal windows from 0.25 to 1.25 s after each indicated perceptual switch, excluding switches that were followed by another switch within 2 s. This relatively short window was tenable as all subjects reported the percept to be most salient right after the perceptual switch. Epochs with the electro-oculogram exceeding 150 μV within the analysis window were rejected, and the accepted epochs were grouped according to the reported percept.

The instantaneous power of MEG and cortical signals was estimated by convolving the raw signals with complex Morlet wavelets (12 cycles) in 4-s windows centered around each perceptual switch and by averaging the squared magnitudes of the convolutions across all switches.

A general linear model (GLM), which included regressors for both tag signals (12 Hz and 15 Hz), subject-specific spontaneous oscillations (9–11 Hz and 17–19 Hz) (31), line frequency interference (50 Hz), and a linear trend, was fitted to the signals in the 1-s windows at every source point on the cortex (see Fig. 4). Both sine and cosine terms were included in each oscillatory regressor to accommodate the unknown phase of the signal. This model was used to obtain the tag-related amplitudes both from the average and continuous MEG data.

The subject-specific statistical parametric maps of the averaged data were converted to binary maps by assigning a ''1'' to all cortical locations where the tag-related signal amplitude exceeded the SD of the background noise by a factor of 4 (corresponding to $P = 0.0051$, Bonferroni-corrected for multiple comparisons) and a ''0'' to the others. The binary maps were then morphed into a common space (''fsaverage'' average cortical surface provided in the FreeSurfer software) and averaged (32) for group-level analysis. An ROI defined on the average cortical surface was then morphed to each individual brain to obtain the mean amplitudes of the tag-related signals within the ROI.

The amplitude ratio (vase tag divided by face tag) was evaluated for signals originating within the ROI and tested for a statistically significant change between the percepts with a binomial test assuming equal probabilities for an increase and decrease.

1. Kleinschmidt A, Büchel C, Zeki S, Frackowiak RS (1998) Human brain activity during spontaneously reversing perception of ambiguous figures. *Proc Biol Sci* 265:2427–2433.
2. Tong F, Meng M, Blake R (2006) Neural bases of binocular rivalry. *Trends Cogn Sci* 10:502–511.
3. Leopold DA, Logothetis NK (1996) Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature* 379:549–553.
4. Tononi G, Srinivasan R, Russell DP, Edelman GM (1998) Investigating neural correlates of conscious perception by frequency-tagged neuromagnetic responses. *Proc Natl Acad Sci USA* 95:3198–3203.
5. Srinivasan R, Petrovic S (2006) MEG phase follows conscious perception during binocular rivalry induced by visual stream segregation. *Cereb Cortex* 16:597–608.
6. Brown RJ, Norcia AM (1997) A method for investigating binocular rivalry in real-time with the steady-state VEP. *Vision Res* 37:2401–2408.
7. Lansing RW (1964) Electroencephalographic correlates of binocular rivalry in man. *Science* 146:1325–1327.
8. Uusitalo MA, Williamson SJ, Seppä MT (1996) Dynamical organisation of the human visual system revealed by lifetimes of activation traces. *Neurosci Lett* 213:149–152.

9. De Marco A, Penengo P, Trabucco A (1977) Stochastic models and fluctuations in reversal time of ambiguous figures. *Perception* 6:645–656.
10. Appelbaum LG, Wade AR, Vildavski VY, Pettet MW, Norcia AM (2006) Cue-invariant networks for figure and background processing in human visual cortex. *J Neurosci* 26:11695–11708.
11. Sterzer P, Kleinschmidt A (2007) A neural basis for inference in perceptual ambiguity. *Proc Natl Acad Sci USA* 104:323–328.
12. Pitts MA, Gavin WJ, Nerger JL (2008) Early top-down influences on bistable perception revealed by event-related potentials. *Brain Cogn* 67:11–24.
13. Leopold DA, Logothetis NK (1999) Multistable phenomena: changing views in perception. *Trends Cogn Sci* 3:254–264.
14. Lumer ED, Friston KJ, Rees G (1998) Neural correlates of perceptual rivalry in the human brain. *Science* 280:1930–1934.
15. Windmann S, Wehrmann M, Calabrese P, Güntürkün O (2006) Role of the prefrontal cortex in attentional control over bistable vision. *J Cogn Neurosci* 18:456–471.
16. Hesselmann G, Kell CA, Eger E, Kleinschmidt A (2008) Spontaneous local variations in ongoing neural activity bias perceptual decisions. *Proc Natl Acad Sci USA* 105:10984–10989.

Parkkonen *et al.*

17. Fang F, Kersten D, Murray SO (2008) Perceptual grouping and inverse fMRI activity patterns in human visual cortex. *J Vis* 8:2.2–9.

18. Murray SO, Boyaci H, Kersten D (2006) The representation of perceived angular size in human primary visual cortex. *Nat Neurosci* 9:429–434.

19. Polonsky A, Blake R, Braun J, Heeger DJ (2000) Neuronal activity in human primary visual cortex correlates with perception during binocular rivalry. *Nat Neurosci* 3:1153–1159.

20. Haynes J, Rees G (2005) Predicting the stream of consciousness from activity in human visual cortex. *Curr Biol* 15:1301–1307.

21. Maier A, Logothetis NK, Leopold DA (2007) Context-dependent perceptual modulation of single neurons in primate visual cortex. *Proc Natl Acad Sci USA* 104:5620–5625.

22. Lee TS, Yang CF, Romero RD, Mumford D (2002) Neural activity in early visual cortex reflects behavioral experience and higher-order perceptual saliency. *Nat Neurosci* 5:589–597.

23. Lee TS, Mumford D, Romero R, Lamme VA (1998) The role of the primary visual cortex in higher level vision. *Vision Res* 38:2429–2454.

24. Lee TS, Mumford D (2003) Hierarchical Bayesian inference in the visual cortex. *J Opt Soc Am A Opt Image Sci Vis* 20:1434–1448.

25. Likova LT, Tyler CW (2008) Occipital network for figure/ground organization. *Exp Brain Res* 189:257–267.

26. Taulu S, Kajola M, Simola J (2004) Suppression of interference and artifacts by the Signal Space Separation method. *Brain Topogr* 16:269–275.

27. Fischl B, Sereno MI, Dale AM (1999) Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage* 9:195–207.

28. Fischl B, Liu A, Dale AM (2001) Automated manifold surgery: constructing geometrically accurate and topologically correct models of the human cerebral cortex. *IEEE Trans Med Imaging* 20:70–80.

29. Dale AM *et al.* (2000) Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron* 26:55–67.

30. Lin F, Belliveau JW, Dale AM, Hämäläinen M (2006) Distributed current estimates using cortical orientation constraints. *Hum Brain Mapp* 27:1–13.

31. Hari R, Salmelin R (1997) Human cortical oscillations: a neuromagnetic view through the skull. *Trends Neurosci* 20:44–49.

32. Fischl B, Sereno MI, Tootell RB, Dale AM (1999) High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum Brain Mapp* 8:272–284.