Stereomotion processing in the non-human primate brain

Yseult HEJJA-BRICHARD[1,2], Samy RIMA[1,2], Emilie RAPHA[1,2], Jean-Baptiste DURAND[1,2], Benoit R. COTTEREAU[1,2]

[1] Centre de Recherche Cerveau et Cognition, Université de Toulouse, 31052 Toulouse, France.

[2] Centre National de la Recherche Scientifique, 31055, Toulouse, France.

Abbreviated title: Stereomotion processing in the macaque brain

Corresponding authors: Yseult Héjja-Brichard (yseult.hejja@cnrs.fr) & Benoit R Cottereau (benoit.cottereau@cnrs.fr)

Number of pages: 37

Number of figures: 8

Number of tables: 1

Number of words: Abstract (250), Introduction (461), Discussion (1,326)

Conflict of interest: None

Acknowledgments: This research was supported by the Agence Nationale de la Recherche Jeunes Chercheuses et Jeunes Chercheurs (Grants ANR-16-CE37-0002-01, 3D3M) and an ATP funding from the Paul Sabatier University awarded to B.R.C.

Abstract

The cortical network that processes cyclopean stereomotion (CSM) was characterized with functional magnetic resonance imaging in 2 awake macaques. The experimental protocol was similar to previous human neuroimaging studies. We contrasted the BOLD responses evoked by dynamic random-dot patterns with backward/forward stereo-3D-motion with those to a control condition that shared the same monocular and binocular properties, except that the temporal frames were shuffled. Voxel-wise analysis shows that in the two hemispheres of each animal, 3 cortical areas exhibit statistically significant sensitivity to CSM. The first area was localized in the fundus of the superior temporal sulcus (CSM_{STS}). The two other areas and were respectively found on the lower bank of the infero-temporal gyrus (CSM_{ITG}) and in the caudal part of the post-parietal sulcus (CSM_{PPC}). Additional retinotopic mapping showed that CSM_{STS} is directly anterior to the MT cluster (V4t, MT, MSTv, and FST) but does not overlap with it. To complete our analyses, we performed analyses within retinotopic areas and found that specific responses to CSM existed within the MT cluster (most notably in areas MSTv and FST) but were not as pronounced as in the 3 CSM areas. Finally, we showed using an additional localizer that areas CSM_{STS} and CSM_{ITG} (but not CSM_{PPC}) also respond to monocular motion and are therefore not uniquely selective to stereomotion. Altogether, our results are in agreement with previous findings in both humans and macaques and suggest that the cortical networks that process cyclopean stereomotion in the two species might be at least partially homologous.

Significance statement

Motion perception is a fundamental property of the visual system in most animal species. Although numerous studies examined how the primate brain processes 2D motion, much less is known about how it encodes 3D motion. A few neuroimaging investigations in human recently established that stereomotion is notably encoded within the hMT+ complex and its neighbourhood. Here, we extended this work to non-human primate using fMRI in 2 awake, behaving macaques. We found stereomotion selective areas in the Superior Temporal Sulcus (STS), in the Infero-Temporal Gyrus (ITG), and in the Post-Parietal Sulcus (PPC). Our data suggest that stereomotion might be at least partially processed by similar neural mechanisms in human and macaque.

Introduction

Motion perception is a fundamental property of the visual system in most animal species. It enables to track over time the position of elements in a scene and thereby facilitates navigation or interactions with moving objects. Numerous studies have characterised planar motion processing in the primate nervous system. In macaque, single-cell recordings showed that it is computed at the cortical level within a specific cortical network that begins in the primary visual cortex and includes higher-level cortical areas, notably located within the superior temporal sulcus (STS) where area MT hosts neurons whose responses are highly selective to motion direction (see e.g. Maunsell & Newsome, 1987) and also reflects motion perception (Newsome & Paré, 1988; Britten et al., 1996). In human, neuroimaging studies suggested that planar motion is also processed within an extended network that includes a putative homologue of area MT: hMT (Huk et al., 2002). Over the last 20 years, the emergence of monkey fMRI has led to further characterise the correspondence between the networks involved in motion processing in the two

species (Vanduffel et al., 2001; Orban et al., 2003) and to confirm that area MT but also its satellite areas (V4t, FST, and MSTv) in macaque have responses to planar motion that are comparable to those of area hMT and its satellites (pV4t, pFST, and pMSTv) in human (Kolster et al., 2009; 2010).

Rather surprisingly, much less is known about the cortical networks that process motion in depth in primate despite being a very common form of motion in everyday life that can notably signal approaching objects. In human, neuroimaging studies based on analyses within regions of interest (ROIs) showed that the hMT+ complex had significant responses to the two binocular cues for motion in depth: the change of disparity over time (CDOT) and the inter-ocular velocity difference (IOVD) (Rokers et al., 2009; Joo et al., 2016). Another fMRI study suggested that the strongest responses to CDOT could come from a cortical region anterior to the hMT+ complex (Likova & Tyler, 2007). In macaque, recent electrophysiological recordings demonstrated that area MT is selective to motion in depth (Czuba et al., 2014) but that this selectivity is primarily driven by the IOVD cue, with only a small contribution from the CDOT cue (Sanada & DeAngelis, 2014). It is therefore possible that in macaque CDOT is also processed in a region anterior to MT and its satellites.

In the present study, we used fMRI recordings in macaque to determine the cortical regions that have specific responses to cyclopean stereomotion (CSM) based on changing disparity over time (CDOT). We used an experimental protocol that was directly adapted from the ones used to investigate CSM processing in humans, as described in Likova and Tyler (2007) and in Rokers, Cormack, and Huk (2009).

Materials and Methods

Subjects

Two female rhesus macaques (age: 15-17 years; weight: 5.35-6.15 kg) were involved in the study. Animal housing, handling, and all experimental protocols (surgery, behavioural training, and MRI recordings) followed the guidelines of the European Union legislation (2010/63/UE) and of the French Ministry of Agriculture (décret 2013–118). All projects were approved by a local ethics committee (CNREEA code: C2EA – 14) and received authorisation from the French Ministry of Research (MP/03/34/10/09). Details about the macaques' surgical preparation and behavioural training are provided elsewhere (see for instance Cottereau et al., 2017).

Data Availability

Data and analysis code will be made available after acceptance of the paper on dedicated platforms (PRIME-DE and OSF).

Experimental design

Our stimuli were directly derived from those of a previous fMRI study that investigated how cyclopean stereomotion (motion-in-depth based on CDOT) is processed in human (Likova & Tyler, 2007, see also Rokers et al. (2009) or Kaestner et al., *In press*). Our aim was to facilitate the comparison between the cortical networks involved in the two species. We used dynamic random-dot stereograms (dRDS) located within a disk (11 degrees of radius) and refreshed at 30Hz. The dot coverage was 15%. To manipulate binocular disparity between the two retinal projections, dots were green in one eye and red in the other and stimuli

were observed through red-green anaglyphs (stimulus code made available on OSF: https://osf.io/yxrsv/). In the ‘cyclopean stereomotion’ (‘CSM’) condition, dots within the upper and lower parts of the disk changed their disparity in opposition of phase, following a triangular function (1Hz) between +/-23.2744 arcmin (see figure 1-A). This disparity range was chosen so as to maximise the visual system responses to binocular disparities (see e.g. Backus et al., 2001 or Cottereau et al., 2011). The opposition of phase between stereomotion of the dots in the upper and lower parts of the disc allowed to get a null and constant average disparity across the visual field at each frame and thereby to prevent stimulus-induced change in vergence eye movement. Note that in this condition, motion in depth is defined from the change of disparity over time (CDOT). We chose to use the term cyclopean stereomotion in reference to the original human fMRI study of Likova and Tyler (2007).

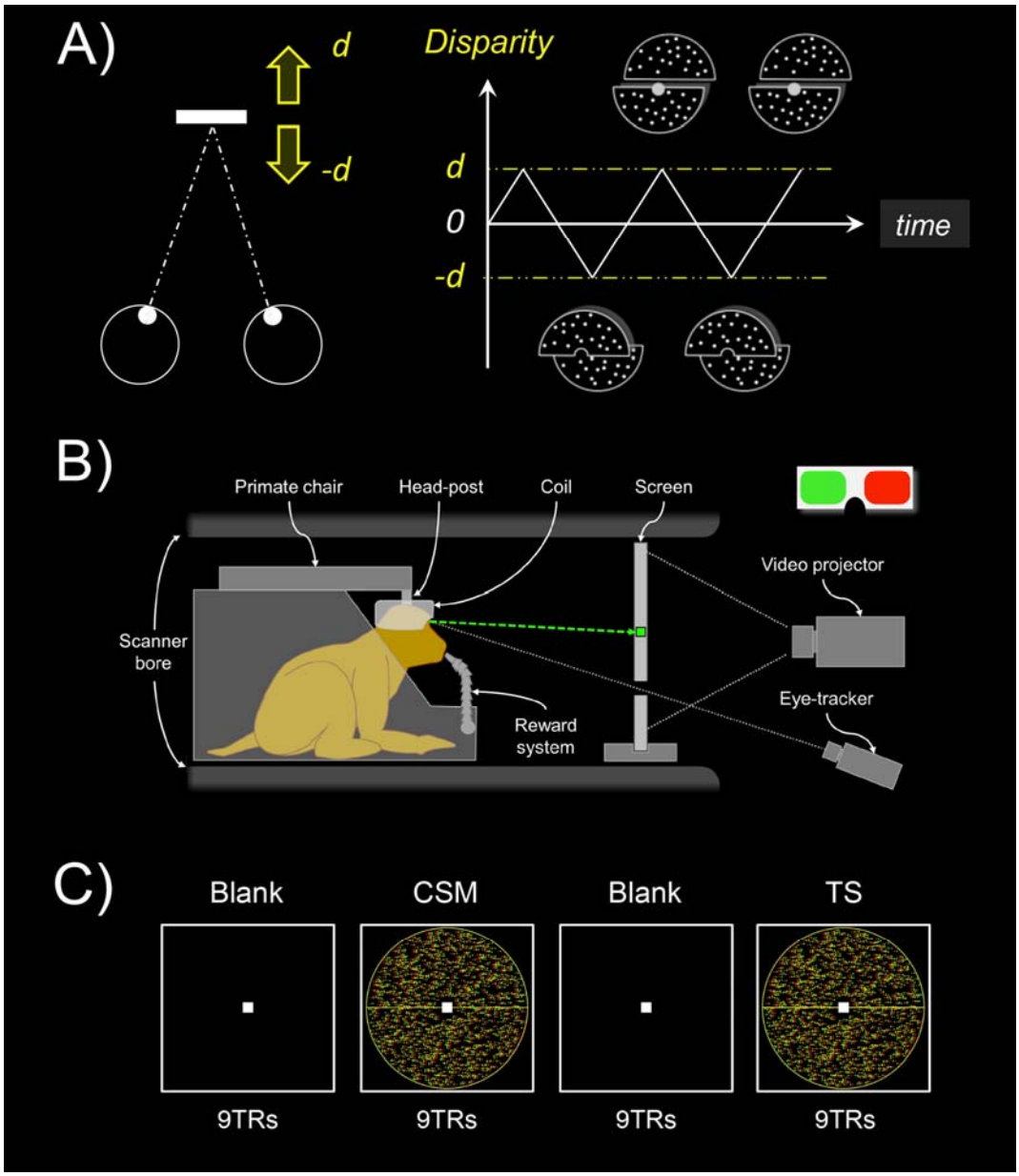

Figure 1: Stimulus design and experimental protocol. A) In the main condition ('CSM' for cyclopean stereomotion), the motion occurs along the antero-posterior axis (leftward panel). The stimulus consisted of a disk defined by dynamic random dot stereograms (dRDS) refreshed at 30Hz. Its upper and lower parts moved in depth between d = ±23.27arcmin in opposition of phase, following a 1Hz triangular function (rightward panel). In our control condition ('TS' for temporally scrambled), the individual frames of the CSM condition were shuffled in time to disrupt the smooth change of disparity over time. Our two conditions had identical retinal disparity distributions but only the CSM condition conveyed motion-in-depth. B) Schematic representation of the monkey fMRI setup. The animal was seated in a sphinx position within the primate chair, in the bore of the scanner, with the 8-channel, phased-array coil placed on top of the head. The animal was involved in a passive fixation task

while its eye position was monitored by an infrared video-based eye-tracker. Horizontal disparity was introduced through the use of red/green anaglyphs. C) Illustration of the experimental protocol. Recordings were performed using a block design, with the alternation of CSM and TS stimuli separated by blank periods. Each run contained 3 repetitions of such blocks plus an additional baseline period of 9TRs (117TRs in total). CSM conditions were shown first in half of the runs and TS conditions appeared first in the other half of the runs.

Our CSM stimulus led to a percept of two planes continuously moving alongside a horizontal axis in opposite directions, one plane being perceived in front of the point of fixation and the other behind the fixation point. The control stimulus consisted of a temporally scrambled version ('TS') of the CSM stimulus. To create this 'temporally scrambled' condition, frames from the 'CSM' stimulus were randomly drawn without replacement. This configuration disrupted the temporal sequence of the stereomotion and thus, the motion in depth. Importantly, both conditions were monocularly identical and contained the same disparity distributions. The average relative disparities between dots in the upper versus lower parts of the disc were also identical between our two conditions.

MRI recordings

Image acquisition: Templates of reference and functional sessions

Whole-brain images were acquired on a 3 Tesla MR scanner (Phillips Achieva) using a custom 8-channel phased-array coil (RapidBiomed) specifically designed to fit the skull of our macaques while preserving their field of view. Four T1-weighted anatomical volumes were acquired prior to the study for each monkey at a high resolution (MPRAGE; repetition time, TR = 10.3 ms; echo time, TE = 4.6 ms, flip angle = 8°; FOV: 155x155 mm; matrix size: 312x192 mm; voxel size = 0.5 x 0.5 x 0.5mm; 192 sagittal slices acquired in an interleaved order), as well as 300

functional volumes (gradient-echo EPI; TR = 2,000 ms, TE = 30 ms, flip angle = 75°, SENSE factor = 1.6; FOV: 100x100 mm; matrix size: 68x64 mm; voxel size = 1.25 × 1.25 × 1.5mm, 32 axial slices acquired in an interleaved order with a thickness of 1.5 mm and no gap). Those data were recorded in a single session whilst the macaques were slightly anaesthetised (Zoletil 100:10 mg/kg and Domitor: 0.04mg/kg) and their constants monitored with an MR compatible oximeter. Those volumes were then used to create individual anatomical and functional templates of reference.

Our T2*-weighted functional images were acquired with a gradient-echo EPI sequence with interleaved slice acquisition (TR = 2,000 ms, TE = 30 ms, flip angle = 75°, SENSE factor = 1.6; FOV: 100x100 mm; matrix size: 68x64 mm; voxel size = 1.25 × 1.25 × 1.5 mm, 32 axial slices acquired in an interleaved order with a thickness of 1.5 mm and no gap).

Scanning Procedure

During the functional recording sessions, macaques were head-fixed and seated in a sphinx position in a dedicated primate chair (see figure 1-B). They had to maintain their gaze within a central fixation window (2x2°) during daily sessions of up to 2 hours. Their fixation was monitored with a monocular ASL© infrared video-based eye tracking setup at 60Hz and they were rewarded through a liquid delivery system at intervals whose frequency depended on their fixation performance. Our stimuli were video-projected on a 768*768 pixel translucent screen (display size: 42.5cm) facing the macaques at a viewing distance of 85cm. We used a block design based on cycles within which our two conditions ('CSM' and 'TS') were interleaved with baseline periods of fixation (see figure 1-C). Both the visual and blank periods lasted

18 seconds (9 TRs) and a cycle was therefore 64-second long (32 TRs). Each run contained 3 repetitions of this cycle plus an extra baseline that was added at the end for a total duration of 117 TR (234 seconds). We displayed the stimuli, controlled for the delivery of the liquid reward and the fixation performance using the EventIDE software (Okazolab®).

Data analysis

Templates of reference

Anatomical and functional templates of reference were created for each individual with the volumes acquired prior to the current study. The anatomical template was obtained with the four T1-weighted anatomical volumes being realigned, averaged, and then co-registered on the MNI space of the 112RM-SL template (McLaren et al., 2009, 2010). To create the functional template, the 300 functional volumes (GE-EPI) were realigned, averaged, and then co-registered on the anatomical template. Mean images were then segmented in order to obtain tissue probability maps for the grey matter, the white matter, and the cerebrospinal fluid (CSF). Spatial normalisation parameters between both templates could then be estimated.

Pre-processing of the raw functional data

Pre-processing and volume-based analyses were carried out with SPM12 in the Matlab® environment. Only runs with central gaze fixation above 85% were kept for further analysis. In total, we kept 43 and 60 runs for both macaques, respectively. The 4 first volumes of each run were discarded (dummy scans) to keep only the BOLD steady-state response. Pre-processing was performed for each volume, run by run. Slice-timing correction (reference slice: 31st slice – middle one) was

performed first. Images were then reoriented, co-registered with the EPI template, and deformed to fit the individual T1 template. No motion correction was applied to the images. Finally, the images were smoothed with a spatial Gaussian kernel (FWHM = 2x2x2 mm).

HRF estimation

Prior to our statistical analyses, we used independent datasets to characterise the BOLD haemodynamic impulse response functions (HRF) separately for each animal. These datasets respectively contained 16 (M01) and 12 (M02) 204s long runs that consisted of 6 cycles of 4s full field counter phasing (10Hz) checkerboards separated by a 30s blank interval (see more details about this procedure in Cottereau et al., 2017). Data were pre-processed using the pipeline described above and projected onto individual surfaces generated with the CARET software (Van Essen et al., 2001). Following Dumoulin and Wandell's procedure (2008), we extracted the BOLD responses from nodes within the anatomically defined V1 of each individual. We only kept visually responsive nodes, that is those whose signal-to-noise ratio (SNR) was greater than 3. This SNR was estimated with a Fourier analysis of the average time courses across runs where the signal corresponded to the Fourier coefficient amplitude at the stimulation frequency F (i.e. F = 1/34) and the noise was given by the average moduli at the two neighbouring frequencies (i.e. $F - \delta f$ and $F + \delta f$, where $\delta f = 1/2$ is the resolution of our frequency analysis). We computed the average time course of these nodes during one cycle and used this average time course for estimating the HRF. The HRF was derived as the response to a 2s stimulus (our fMRI sampling rate). Note however that our stimulus duration was 4s rather than 2s because linearity deteriorates at short durations (Boynton et al. 1996; Logothetis and Wandell, 2004) and also because this duration was used in

a previous monkey fMRI study that characterised the BOLD HRF in macaque (Leite et al., 2002). For each monkey, the average response to the 4s stimulus was fit as the convolution of the responses to two 2s responses, each of which is the HRF. We parameterised the HRF as the difference of two gamma functions (Friston et al., 1998). This functional form of the HRF captures the late undershoot of the response better than a single gamma function (Boynton et al., 1996).

General linear model (GLM)

Univariate statistics were performed at the voxel level, using a general linear model (GLM). Our visual (CSM, TS) and baseline conditions were implemented as the 3 main regressors of the GLM. As reported above, we only analysed runs with fixation performance greater than 85%. We also made sure that there was no difference in fixation stability between our two visual conditions during these runs. This was done using paired t-tests that did not show any significant effect (see figure 2). Finally, we used the oculometric data of those runs to define regressors of non-interest that were included in the GLM so as to exclude the possible contribution of eye movement from our analyses. These regressors were obtained by detecting the presence (1) or absence (0) of saccades in the different volumes of every run. Given the limited resolution of our oculometric data, we chose a loose threshold to detect saccades, considering a saccade to be present when the eye signal was distant from more than the median ±5 mad (median absolute deviation) from the central fixation point. The corresponding saccade regressors were then convolved with the HRF and introduced into the model. To characterise and eliminate noise in our recordings, we also performed a principal component analysis on voxels located outside the brain (see Farivar and Vanduffel, 2014). Indeed, time courses in those voxels should not be influenced by our experimental design but rather reflect

physiological noise and artefacts caused by movements of the animal. For each run, we determined the number of principal components that were necessary to explain 80% of the variance in these voxels and used the corresponding principal vectors as regressors of non-interest in our model. This adaptive procedure typically added an average of 13.2553 (±9.2653) and 11.2791 (±5.0584) additional regressors in the model.

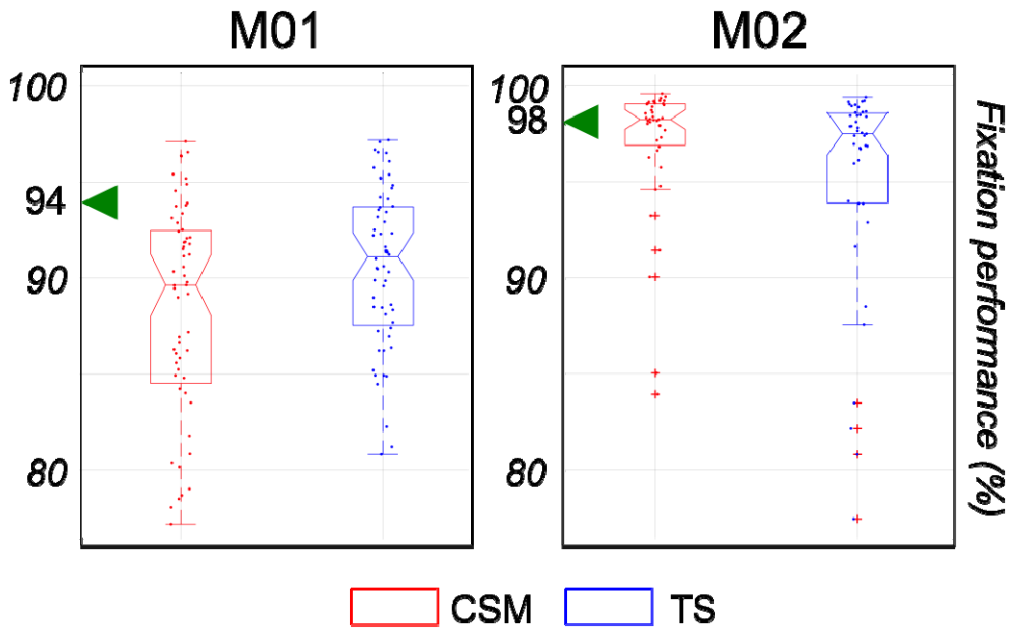

Figure 2: Ocular behaviour. Boxplots represent the distribution of percentages of correct fixation (i.e. less than 2° away from the fixation point) for both CSM and TS conditions and for each monkey (median and interquartile values). Numbers and arrows in green indicate the median values of the percentage of fixation performance during the baseline condition. Jittered dots are the fixation performance values for every run. Fixation performances were not different between our two conditions (paired t-tests, p>0.3 and confidence intervals including 0).

We estimated the beta values associated with our GLM using the RobustWLS toolbox (Diedrichsen & Shadmehr, 2005), which is provided as an additional toolbox for SPM12 (http://www.diedrichsenlab.org/imaging/robustWLS.html). This approach allows estimating the noise variance for each image in the time series, using the

derivative of a maximum likelihood algorithm. Variance parameters are then used to obtain a weighted least square estimate of the regression parameters of the GLM. It therefore helps reduce the impact of noisier volumes on beta estimation. Previous studies showed that such a method significantly improved estimations in block design fMRI experiments (see e.g. Takeuchi et al., 2011).

Statistical analyses

As a first step for identifying potential areas with specific responses to motion-in-depth, we used SPM12 to compute the statistical parametric map corresponding to the contrast between our two visual conditions ('CSM' and 'TS') and thresholded this map at $p<10^{-3}$ (uncorrected, t-value>3.1). Cortical regions that showed significantly stronger responses to CSM than to TS in both hemispheres and in each animal were considered as ROIs. In a second step, we followed a cross-validation procedure to characterise the difference in the percentage of signal change (PSC) between our two conditions in these ROIs (see Kriegeskorte et al., 2009). In this procedure, we recomputed the statistical parametric map corresponding to the contrast between CSM and TS from a subset of the data made of all the runs but one. We determined the local maxima corresponding to our ROIs in this subset by searching within 5x5x5 voxel cubes centred on the local maxima obtained using the whole dataset. Once the coordinates of these subset local maxima were identified, we considered 3x3x3 voxel cubes around these coordinates and we computed the PSC differences between our two conditions in the left-out run from the beta weights extracted in these voxel cubes. This procedure was repeated by successively leaving out each of the runs, respectively resulting in 60 and 43 values for each ROI in M01 and M02. This method avoids double dipping because the data that were used to define the local maxima are independent of those that were used to

characterise the activity (Kriegeskorte et al., 2009). To test the robustness of that method, we repeated the operation with smaller (3x3x3) or larger (9x9x9) search areas, which did not lead to different results. Finally, our beta values were extracted within small (3x3x3) voxel cubes rather than within patches determined by anatomical and/or statistical considerations, because anatomical borders between areas are difficult to determine precisely and our contrast led to extended activations that cannot be accurately divided into clusters corresponding to different functional regions (see figures 3 and 4). Our approach is more conservative and avoids subjectivity when dealing with borders between areas. Importantly, we reproduced our analyses with beta extracted from smaller (1x1x1) or larger (5x5x5) voxel cubes, which did not impact our results.

We used the following equation to compute the difference of PSC between our two conditions:

$$\Delta_{PSC} = PSC_{CSM} - PSC_{TS},$$

with:

$$PSC_{CSM} = (\beta_{CSM} - \beta_{baseline}) / \beta_{constant} \times 100$$
$$PSC_{TS} = (\beta_{TS} - \beta_{baseline}) / \beta_{constant} \times 100$$

Note that we chose to use the difference of PSCs because the PSCs for the CSM and TS conditions are paired.

In order to estimate whether our observed PSC differences were not due to chance, we computed permutation tests. We randomly attributed a negative sign to our PSC values and computed the mean difference, repeating this procedure 10,000 times. We then calculated a *p*-value defined as the proportion of random differences that were superior to our observed difference.

ROI analysis

We also performed a wide-field retinotopic mapping to delineate retinotopic regions that were used for additional ROI-based analyses. Whole-brain images were acquired with an identical setup as for the main experiment. Visual stimuli were displayed using a large field-of-view (80° of visual angle, viewing distance = 25cm) and consisted of videos of a fruit basket that was lateral as well as forward and backward monocular motions. Traditional (clockwise/counter clockwise) rotating wedges (radius: 40°, angular extent: 49°) and expanding/contracting rings (eccentricity linearly varying between 0° and 40°) were used as visual apertures through which the fruit basket was displayed. Each run lasted 230s and contained 5 cycles of 44s with the first 10 seconds of a run being discarded (dummy scans) for the signal to reach its baseline. A small green square (0.4° x 0.4°) at the centre of the screen was used to control for fixation during passive viewing. As in our main experiment, only runs with more than 85% of correct fixation (respectively 47 and 48 runs for M01 and M02) were kept for further analyses. A pre-processing pipeline similar to the one described above was performed on the selected runs except that no smoothing was applied to the volumes and a fixed number of components (18 components) was used when performing the PCA, later used as a regressor of non-interest in the GLM. We projected the volume data onto individual surfaces using the Caret software (Van Essen et al., 2001) and a custom reorientation algorithm.

A population receptive field (pRF) analysis (Dumoulin & Wandell, 2008) was performed using the Matlab analyzePRF toolbox developed by Kay et al., (2013). For each surface node, an exhaustive set of theoretical pRF parameters (polar angle, eccentricity and size) was used to generate time courses that were compared

to the real recordings. pRF size and position parameters that best predicted the data were selected as the node pRF parameters. With this approach, we obtained polar angle and eccentricity maps from which we characterised retinotopic areas that were described in previous monkey fMRI studies: V1, V2, V3, V4, as well as the regions within the STS (V4t, MT, MSTv, and FST) that form the MT cluster as described by Kolster et al. (2009). Those 8 retinotopically-defined regions were then projected back to the volumetric space to perform a ROI-based analysis of our motion-in-depth data. This was done using the inverse of the transformation between the volumetric and surface spaces mentioned above.

Monocular motion localisers

To determine whether the regions that have specific responses to 3D motion are also responsive to monocular motion, we performed a control experiment in which we contrasted responses to static images versus rich monocular motion stimuli. The scanning procedure was identical to the main experiment procedure. Motion localiser stimuli were based on the fruit basket video used for the retinotopic mapping experiment. For the static version, static images were randomly drawn from the video and refreshed at 1Hz. For the moving version, the video was normally played. Stimuli were displayed either centrally (<3° of eccentricity) or peripherally (>3° of eccentricity) and were projected on a 700x700 pixel translucent screen (display size: 42.5cm) facing the macaques at a viewing distance of 25 cm. Each visual condition lasted 6 seconds and was interleaved with a 10-second baseline. The four visual conditions were presented in a pseudo-randomised order and were repeated 3 times within each run. Five extra baseline scans were added at the beginning of a trial for the signal to reach its baseline, thus resulting in a total duration of 202 seconds for each run. In total 42 and 26 runs with fixation above

85% were kept for our analyses. Selected data was pre-processed as previously described, with an adaptive number of components that were necessary to explain 80% of the variance for each run, adding an average of 12.5714 (±10.0369) and 11.9615 (±3.6930) additional regressors in the model.

To estimate motion sensitivity in our regions of interest and in our retinotopic areas, we contrasted moving conditions and static conditions, by gathering central and peripheral simulations. We then performed a ROI-based analysis, looking at the BOLD activity within our independently defined regions.

Results

The aim of this study was to identify the cortical networks that process motion-in-depth defined by the change of disparity over time in macaque, called here cyclopean stereomotion (CSM). Our experimental design was directly derived from previous human fMRI studies (Likova & Tyler, 2007; Rokers et al., 2009) so as to determine the homologies but also the differences between the BOLD activations in the two species (Orban, 2002). Our CSM condition and its temporal scramble ('TS') control were defined from dynamic random dots stereograms (dRDS). They had identical retinal disparity distributions but differ in their temporal structure (see the materials and methods section). Only the CSM condition conveyed motion-in-depth. Figures 3 and 4 show the statistical parametric maps obtained for the contrast between 'CSM' and 'TS' on the individual anatomical templates of each subject (M01 on figure 3 and MO2 on figure 4). These data are shown for different coronal slices.

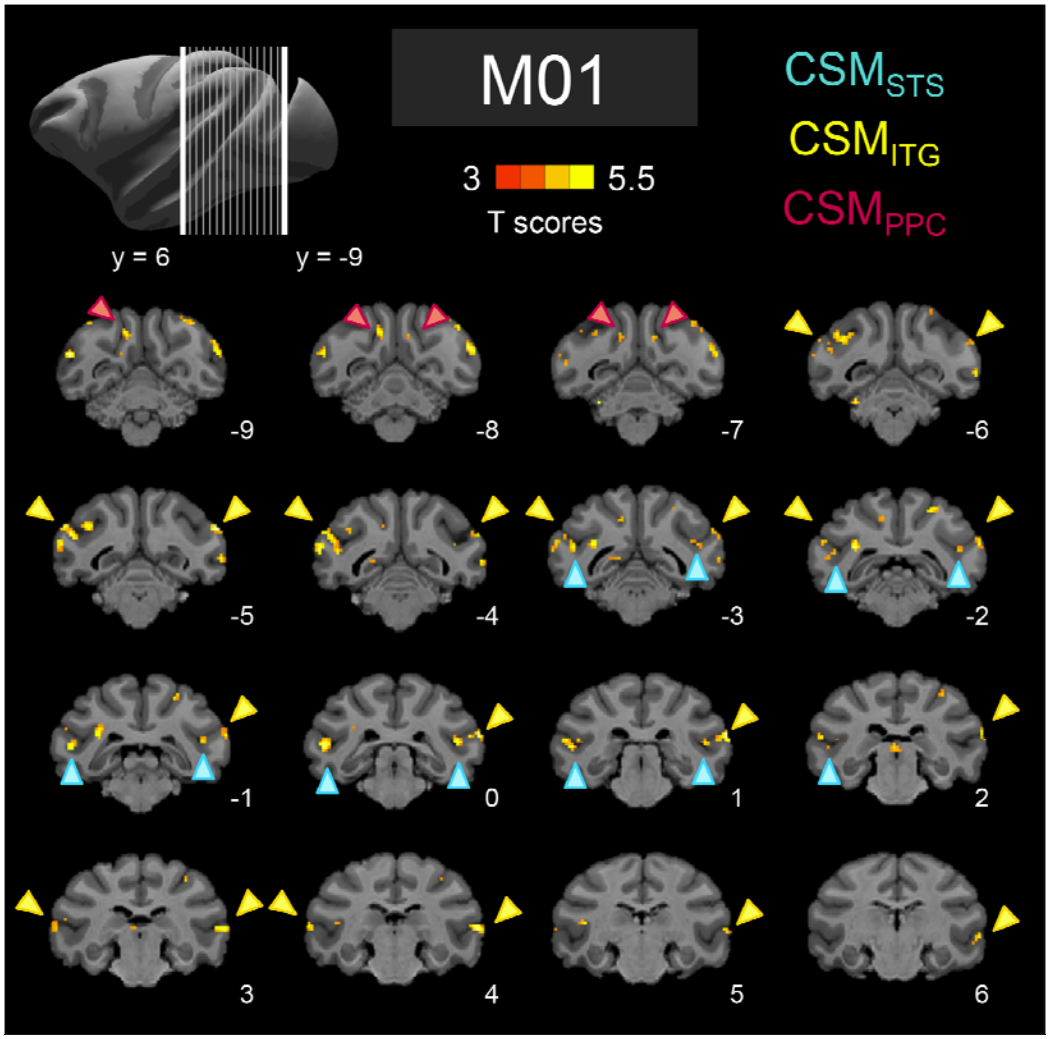

Figure 3: Activations for the contrast between Cyclopean Stereomotion (CSM) and its temporally scrambled version (TS) for M01.Figure shows activations that were stronger for the CSM condition than for the TS condition. Data are projected on the individual anatomical template of the macaque and are shown for different coronal slices. Coloured arrows indicate the localisation of our three regions of interest: CSM_{STS} (in blue), CSM_{ITG} (in yellow), and CSM_{PPC} (in pink). T-scores were obtained after computing the statistical parametric map for the contrast of interest between CSM and TS.

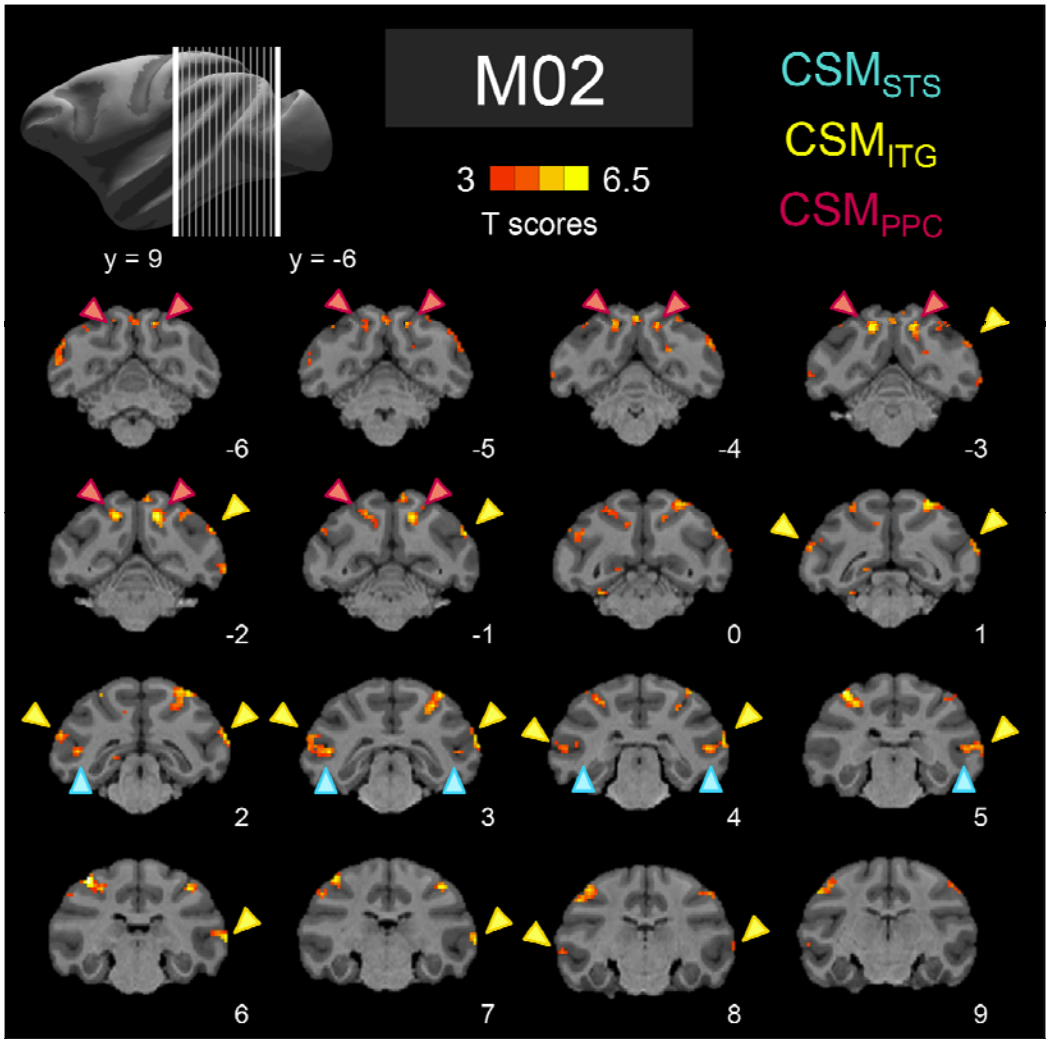

Figure 4: Activations for the contrast between Cyclopean Stereomotion (CSM) and its temporally scrambled version (TS) for M02. See figure 3 for the details of the legend.

On figures 3 and 4, red-to-orange colours indicate stronger BOLD activations for the CSM condition ($p < 10^{-3}$, uncorrected). Despite differences in the activation patterns observed in the two animals, this analysis reveals a network encompassing the temporal and parietal cortices in both monkeys. Three areas of this network are consistently found in the two hemispheres of the two macaques. Coloured arrows on the figure show these areas. For sake of comparison with previous human neuroimaging studies, we named those areas after Likova and Tyler's denomination (Likova & Tyler, 2007), that is, CSM for cyclopean stereomotion responsive areas.

The first area is located within the superior temporal sulcus (STS), in its fundus (CSM_{STS}). The second one is located on the infero-temporal gyrus (CSM_{ITG}), in the prolongation of the inferior occipital sulcus (IOS). The last area is localised in the posterior parietal cortex (PPC), encompassing the intra-parietal sulcus (CSM_{PPC}). The MNI coordinates corresponding to the local maxima of these areas in the two animals are provided in table 1.

It is worth noting that in monkey M02, significant BOLD activations were also found in more anterior parts of the IPS (*t*-score > 3.1), notably within the ventral intraparietal area (VIP) and the antero-intraparietal area (AIP). VIP has been shown to play a role in egomotion processing in both monkey (Cottereau et al., 2017) and human (Wall & Smith, 2008), whereas AIP has been suggested to play a central role in depth perception and visually guided hand movements in both species as well (Sakata et al., 1997; Shikata et al., 2007). Unfortunately, we were not able to find those activations in the other macaque, potentially because of a slightly smaller SNR.

ROI	**M01**			**M02**		
	x	y	z	x	y	z
CSM_{STS}						
L	-24	0	16	-19	3	16
R	21	0	17	20	4	16
CSM_{ITG}						
L	-29	4	15	-23	3	18
R	28	1	20	26	3	19
CSM_{PPC}						
L	-5	-8	28	-6	-3	30
R	5	-7	27	6	-2	30

Table 1: MNI coordinates (in mm) of the local maxima for the 3 regions that were significantly more responsive for the CSM condition than for the TS control in the two hemispheres of the two animals.

In order to determine the signal variation between our two visual conditions in these 3 regions, we computed the difference between the percentages of signal change (PSC) for CSM versus TS. This was done using a cross-validation procedure where the local maxima were re-estimated from a subset of runs and the corresponding activations were extracted from the remaining part of the data (see the method section for more details). This process avoids double dipping and thereby a biased interpretation of the results (Kriegeskorte et al., 2009). The estimated differences of PSC for our three ROIs are shown in figure 4.

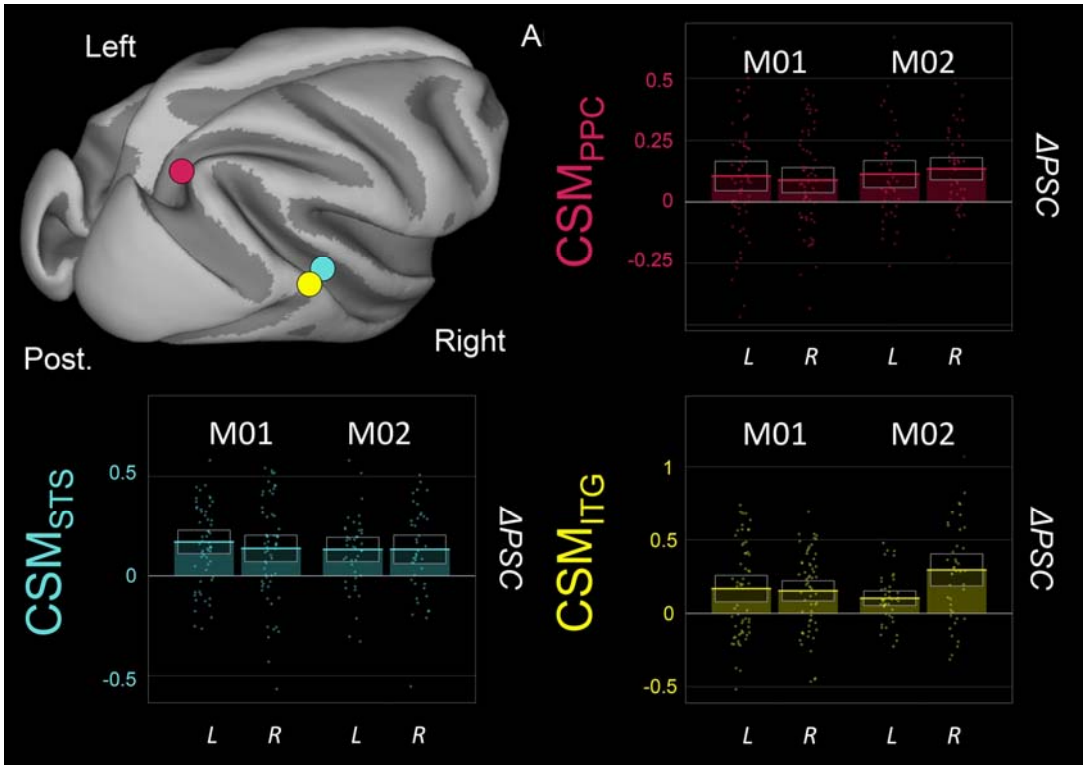

Figure 5: Difference in signal change (ΔPSC) between the CSM and TS conditions in the 3 regions of interest (CSM_{STS}, CSM_{ITG}, and CSM_{PPC}). The thick lines of the bar graphs provide the average values across runs for the left (L) and right (R) hemispheres of the two monkeys (M01 and M02). The boxes give the 95% confidence intervals for these average values. The dots provide the data for each run. A small horizontal jitter was introduced to facilitate visibility.

The thick lines on the bar graphs provide the average values of these differences for the left (L) and right (R) hemispheres in the two macaques (M01 and M02). The boxes give the corresponding 95% confidence intervals. We can observe that for each of our 3 ROIs, these average differences are consistent across hemispheres and animals and also greater than 0, which reflects stronger responses for the CSM condition. This was confirmed by our permutation tests in all 3 regions ($p < 0.001$)

Retinotopic analysis

Previous studies in human found that the hMT+ complex had significant responses to stereomotion, notably based on changing disparity over time (CDOT) (Rokers et al., 2009; Joo et al., 2016). A single-cell study in macaque also found a weak but significant selectivity to CDOT in area MT (Sanada & DeAngelis, 2014). In order to determine whether the CSM-responsive ROIs we obtained from our voxel-wise analyses overlap with (or correspond to) area MT and/or its neighbour regions, we performed a retinotopic mapping in our two animals (see more details in the Materials and methods section). This method notably allowed us to delineate the areas of the MT cluster: V4t, MT, MSTv and FST (see Kolster et al., 2009). On figures 5-A and 6-A, we show the locations of these areas and of our two CSM-responsive regions around the STS and the ITG, CSM_{STS} and CSM_{ITG}, respectively.

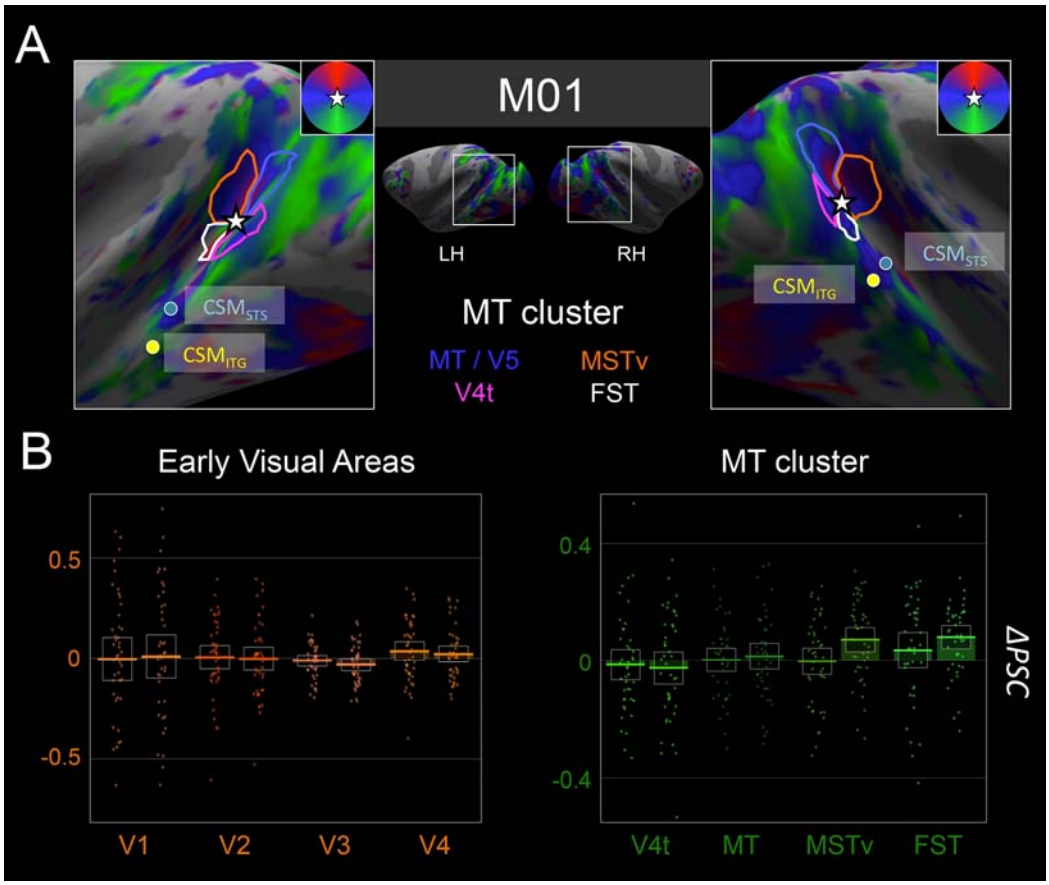

Figure 6: A) Retinotopic mapping of the Superior Temporal Sulcus (STS) for M01 and delimitation of the MT cluster areas: MT (dark blue), V4t (pink), MSTv (orange), and FST (white). We show here the polar angle maps that were used to delineate the borders between these areas. The extent of those areas was obtained from the eccentricity maps. Coloured dots indicate the local maxima positions for areas CSM_{STS} (in blue) and CSM_{ITG} (in yellow). As shown on the maps, CSM_{STS} and CSM_{ITG} are close to but do not overlap with the MT cluster. B) Difference in signal change (ΔPSC) between the CSM and TS conditions in retinotopic areas. On the left, thick lines of the bar graphs provide average values for the left and right hemispheres of early visual areas: V1, V2, V3, and V4. On the right, average values are given for the MT cluster areas: V4t, MT, MSTv, and FST. The boxes give the 95% confidence intervals for the average values. The dots provide the data for each run. A small horizontal jitter was introduced to facilitate visibility.

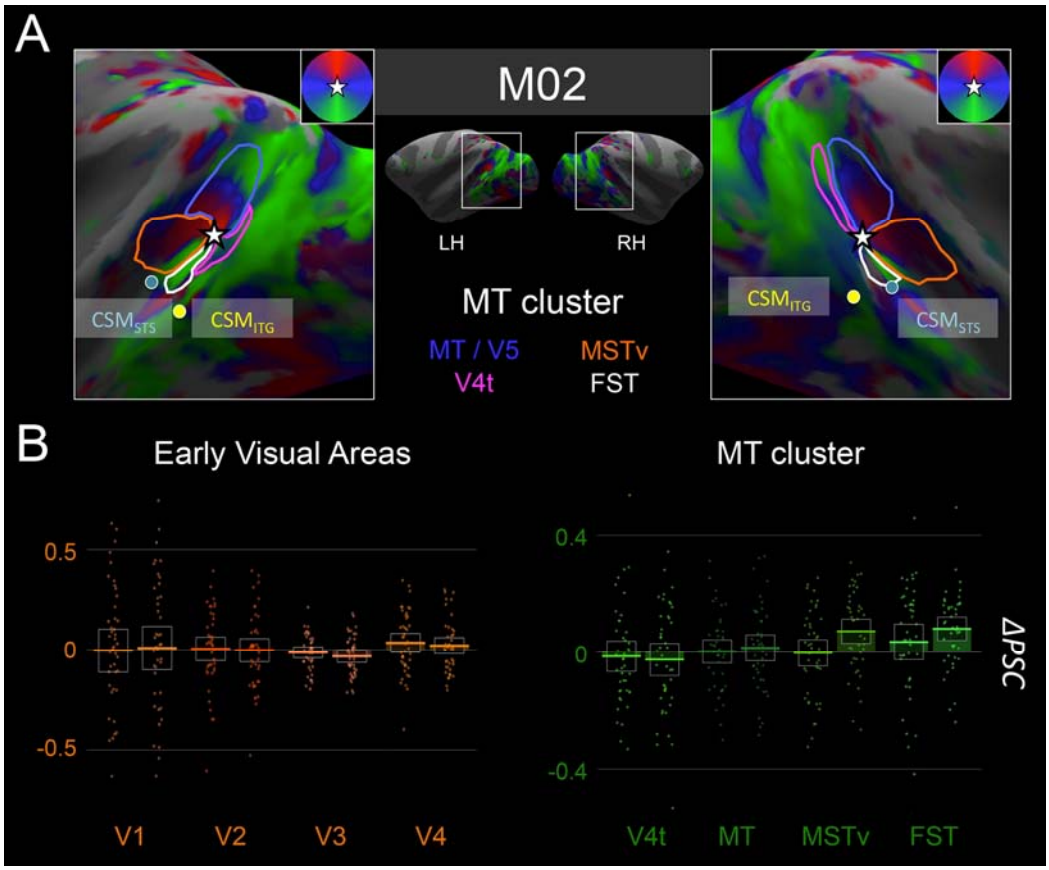

Figure 7: A) Retinotopic mapping of the Superior Temporal Sulcus (STS) for M02 and delimitation of the MT cluster areas: MT (dark blue), V4t (pink), MSTv (orange), and FST (white). See figure 6 for the details of the legend.

We can see that if our two regions are close to the MT cluster, they nonetheless do not overlap with it. Area CSM_{STS} is located at a more anterior position along the STS. Area CSM_{ITG} is located on the ITG next to the STS, in the extension of the intra-occipital sulcus (IOS).

To complete our study and facilitate the comparison with previous findings in human and non-human primates, we also performed ROI-based analyses within the retinotopic regions. The differences between the percentages of signal change (PSC) for CSM versus TS in our two macaques are shown on figure 6-B and 7-B for early visual areas (V1, V2, V3, and V4) and the MT cluster (V4t, MT, MSTv, and FST). We can observe that if CSM selectivity in all these areas is not as pronounced

as in CSM_{STS}, CSM_{ITG}, and CSM_{PPC}, responses in the MT cluster tend to be stronger than those measured in early visual cortex. Permutation tests demonstrated that in the MT cluster, responses to our CSM condition are significantly stronger than to our control condition in areas MT (1/4 hemisphere), MST (2/4 hemispheres, right hemispheres only), and FST (3/4 hemispheres). This suggests that selectivity to cyclopean stereomotion exists in these regions. However, it is not as pronounced as in the areas obtained from our univariate analysis. We did not find any consistent effects in early visual areas.

Monocular motion analysis

To test whether our three regions also had specific responses to monocular motion, we ran an additional motion localiser in our two animals (n = 42 and n = 26 runs in M01 and M02, see more details in the Materials and methods section). We then computed the difference between the percentages of signal change (ΔPSC) corresponding to the monocular motion versus static image conditions. As expected from such a localiser, this analysis led to significantly stronger responses to motion in most of the retinotopic areas and more specifically within areas of the MT cluster. In particular, permutation tests demonstrated that all 4 regions of the MT cluster had significantly stronger responses to monocular motion in the two animals ($p < 0.05$ except for the left V4t in M02). We show in figure 8 the results of these analyses in our 3 CSM responsive areas (CSM_{STS}, CSM_{ITG}, and CSM_{PPC}).

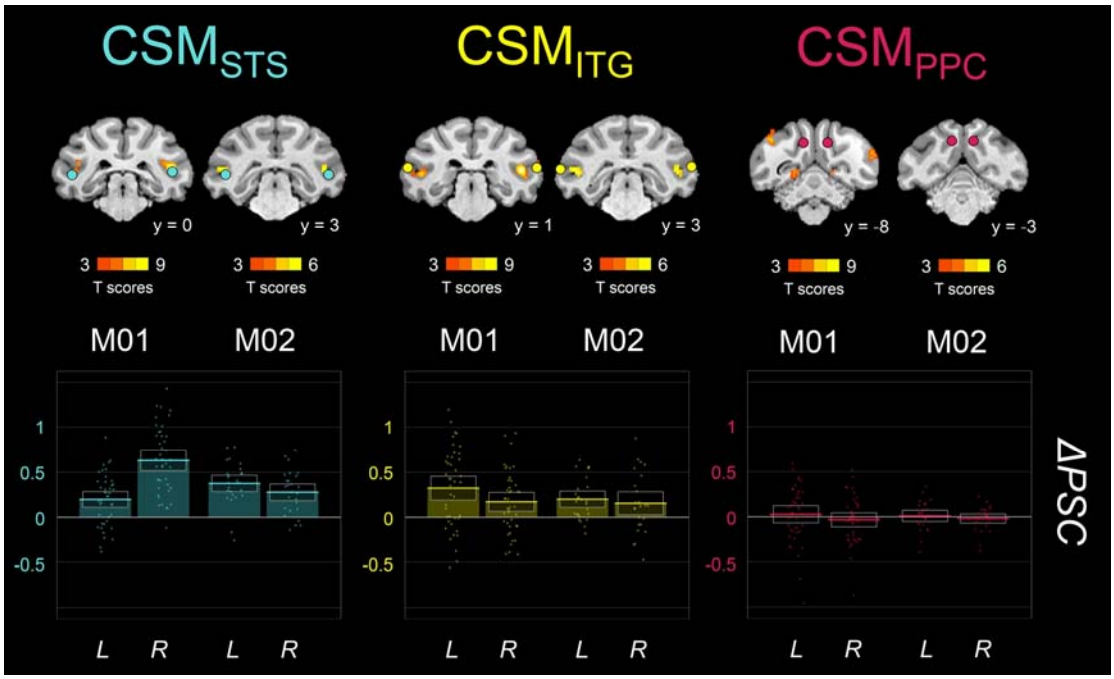

Figure 8: Sensitivity to monocular motion in CSM_{STS}, CSM_{ITG}, and CSM_{PPC}. Strongest responses to monocular motion are shown on coronal slices from the individual anatomical template of each animal (upper panel). The colour dots provide the position of CSM_{STS}, CSM_{ITG}, and CSM_{PPC}*. For these 3 regions, PSC difference between responses to monocular motion vs static image are shown on the graphs of the lower panel. The thick lines of the bar graphs provide the average values across runs for the left (L) and right (R) hemispheres of the two monkeys (M01 and M02). The boxes give the 95% confidence intervals for these average values. The dots provide the data for each run. A small horizontal jitter was introduced to facilitate visibility.*

We can observe that only CSM_{STS} and CSM_{ITG} areas have significant responses to monocular motion (permutation tests, $p < 0.05$), in both hemispheres for CSM_{STS} and in the left hemisphere for CSM_{ITG}, for each monkey. Their motion selectivity (in particular in CSM_{STS}) is therefore not specific to cyclopean stereomotion. On the opposite, responses to monocular motion in area CSM_{PPC} are not different from responses to static patterns (permutation tests, $p > 0.1$). It implies that this region might uniquely respond to cyclopean stereomotion.

Discussion

The aim of the present study was to characterise the cortical networks that process cyclopean stereomotion in macaque monkeys. To that end, we adapted the experimental protocol of two previous human fMRI studies that found significant activations in the hMT+ cluster and/or in an area directly anterior to it (Rokers et al., 2009; Likova & Tyler, 2007). Our main condition ('CSM' for cyclopean stereomotion) and its control ('TS' for 'temporally scrambled') shared the same disparity distribution and were monocularly identical (figure 1) but only the CSM condition conveyed stereomotion, since the temporal sequence was scrambled in the TS condition. We recorded whole-brain BOLD responses from 2 behaving macaques using a block design. Our analyses revealed a network of 3 areas whose responses to our CSM condition were consistently (i.e. across hemispheres and animals) stronger than those to our control condition (figures 3 and 4). In reference to the original study of Likova & Tyler (2007), we labelled those regions CSM_{STS}, CSM_{ITG}, and CSM_{PPC}. To complete these analyses, we also documented responses to our CSM condition in visual areas estimated from independent wide-field retinotopic mapping procedures (figures 6 and 7).

Our CSM_{STS} region is located in the fundus of the superior temporal sulcus, on its inferior bank (figures 3 & 4), at a location (see table 1) anterior to the MT area and its satellites (V4t, MSTv, and FST, see figures 6 and 7). Based on our retinotopic maps, we confirmed that this region is independent from the MT cluster (only a marginal overlap with area FST was found in the right hemisphere of M02) and is more likely directly adjacent to area FST. Our additional motion localiser demonstrated that it is also responsive to monocular motion (figure 8). Previous studies in macaque reported additional motion-sensitive regions in anterior portions

of the STS. Using fMRI, Nelissen et al., (2006) notably documented an area in the lower superior temporal sulcus (LST) that responds to opponent motions and to actions. This area was 6-8mm anterior to FST and therefore does not fully coincide with our CSM_{STS} area. CSM_{STS} might thus be a distinct motion-selective area of the STS with specific responses to 3D motion.

Although univariate statistics did not show significant responses in the MT area and its satellites, ROI-based analyses demonstrated that for some animal and/or hemispheres, responses in MT (1/4 hemisphere), MSTv (2/4 hemispheres), and FST (3/4 hemispheres) were significantly stronger for our stereomotion condition (figures 6 and 7). In a pioneer study, Maunsell and Van Essen (1983) concluded from single-cell recordings in area MT of anaesthetized macaques that neurons in this region had no selectivity to motion-in-depth (see also Felleman and Kaas, 1984). More recently, Sanada & DeAngelis (2014) found, using a different method, that MT does host neurons tuned to motion-in-depth (see also Czuba et al., 2014) but that these neurons were mostly driven by the inter-ocular velocity difference (IOVD) between the two eye with only a modest contribution of the change of disparity over time (CDOT): ~10% of their neurons had significant selectivity for CDOT versus ~57% for IOVD. These findings are in line with our study and suggest that if selectivity to stereomotion is observable in area MT, it remains moderate. To our knowledge, selectivity to motion-in-depth was not directly tested in areas MSTv and FST. Based on our data, we hypothesise that a larger proportion of neurons tuned to cyclopean stereomotion could be found there. Altogether, the responses we measured in the STS are consistent with a model where stereomotion would be progressively integrated along a posterior-to-anterior axis with moderate responses

in MT, intermediate responses in areas MSTv and FST, and stronger responses in CSM_{STS}.

Our stimuli were directly adapted from those used in two human fMRI studies (Likova & Tyler, 2007; Rokers et al., 2009). From a ROI-based analysis, one of these studies found that stereomotion led to significant responses in the hMT+ cluster (Rokers et al., 2009; see also Kaestner et al., *in press*). Using the same analysis, Likova & Tyler (2007) also found significant responses to stereomotion in the hMT+ cluster but their univariate statistics demonstrated that the strongest activations were actually located in a region anterior to hMT+. Our results are in line with both these studies and suggest that our CSM_{STS} region might be the macaque homologue of human area CSM. To clarify this point, it might be interesting for future human studies to properly define area MT and its satellites using retinotopic mapping (see Kolster et al., 2010) in order to determine the exact location of the CSM area with respect to those regions.

Our CSM_{ITG} region is located on the infero-temporal gyrus, next to the STS (figures 3 and 4) and in the direct continuity of the intra-occipital sulcus (IOS). This location matches well with the cluster described in Kolster et al. (2014), which gathers areas V4A, PITd, PITv and OTd. Unfortunately, our retinotopic data did not allow to properly map those regions. Actually, most of these areas have been shown to have strong emphasis on the central visual field, while our retinotopic mapping was intended to cover a large portion of the visual field, with a detrimental effect on the fine-grained definition of foveal regions. Therefore, we cannot clearly determine whether CSM_{ITG} overlaps with one or several of these areas. Compared to CSM_{STS}, responses to monocular motion were only moderate in CSM_{ITG} (figure 8). This is in

line with the findings of Kolster et al. (2014) who reported that motion selectivity in areas V4A, PITd, PITv, and OTd was weaker than in the MT cluster. These regions are selective to 2D shapes and an earlier study suggested that they might also respond to 3D structure defined from motion (Vanduffel et al., 2002). One possibility is that the stereomotion activations in CSM_{ITG} were driven by the structures of the upper and lower half-disks that are more apparent in the CSM condition. If it is the case, it nonetheless constitutes a mechanisms based on stereomotion integration as our two stimuli shared the same (absolute and relative) disparity distributions and only differed in their temporal sequences. Maybe this region is involved in the detection of objects from various cues as for example 3D motion. Further studies will be necessary to clarify its function. Interestingly, in their human study, Rokers et al. (2009) also reported significant responses to cyclopean stereomotion in an area (LO) that is highly responsive to objects.

Area CSM_{PPC} is localized in the caudal part of the intra-parietal sulcus (IPS) within the posterior parietal cortex (PPC). Responses in this region were not stronger for monocular motion than for static stimuli, in agreement with previous monkey fMRI studies (see e.g. Vanduffel et al., 2001). It is known that the caudal part of the IPS is involved in integrating disparity over space, notably for extracting 3D structure (see e.g. Van Dromme et al., 2016). Our results suggest that it is also able to integrate disparity over time to extract 3D motion. This region might therefore play a key role in the detection of moving objects. In human, the studies of Likova & Tyler (2007) and of Rokers et al. (2009) did not explore stereomotion selectivity in the parietal cortex (the latter nonetheless reported significant responses to CDOT in dorsal area V3A). The only study that reported results at the whole-brain level (Kaestner et al., *In press*) found strong stereomotion responses in area IPS-0, which is located in the

caudal part of the human IPS and that is therefore a potential homologue of our CSM_{PPC} region. Further studies will be necessary to clarify this point.

Conclusion

Our fMRI recordings in two macaques demonstrated that cyclopean stereomotion is mainly processed within a network of 3 areas: CSM_{STS}, CSM_{ITG}, and CSM_{PPC}. We also observed a moderate selectivity in areas MT, MSTv, and FST. These results are close to those observed in human using the same experimental protocol and therefore suggest that similar networks in the two species might process stereomotion.

References

- Backus, B.T., Fleet, D.J., Parker, A.J., & Heeger, D.J. (2001). Human Cortical Activity Correlats With Stereoscopic Depth Perception. *Journal of Neurophysiology, 86*(4), 2054-2068.

- Boynton, G.M., Engel, S.A., Glover, G.H., Heeger, D.J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *Journal of Neuroscience, 16*(13), 4207-4221.

- Britten, K. H., Newsome, W. T., Shadlen, M. N., Celebrini, S., & Movshon, J. A. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Visual neuroscience*, *13*(1), 87-100.

- Cottereau, B. R., Smith, A. T., Rima, S., Fize, D., Héjja-Brichard, Y., Renaud, L., Durand, J.-B. (2017). Processing of Egomotion-Consistent Optic Flow in the Rhesus Macaque Cortex. *Cerebral Cortex*, 1-14.

- Cottereau, B. R., McKee, S. P., Ales, J. M., & Norcia, A. M. (2011). Disparity-Tuned Population Responses from Human Visual Cortex. *The Journal of Neuroscience, 31*(3), 954–965.

- Czuba, T. B., Huk, A. C., Cormack, L. K., & Kohn, A. (2014). Area MT encodes three-dimensional motion. *Journal of Neuroscience*, *34*(47), 15522-15533.

- Diedrichsen, J., & Shadmehr, R. (2005). Detecting and adjusting for artifacts in fMRI time series data. *NeuroImage, 27*(3), 624–634.

- Dumoulin, S.O. & Wandell, B.A. (2008). Population receptive field estimates in human visual cortex. *Neuroimage*, *39*, 647– 660.

- Vanduffel, W., & Farivar, R. (2014). Functional MRI of Awake Behaving Macaques Using Standard Equipment. *Advanced Brain Neuroimaging Topics in Health and Disease - Methods and Applications*.

- Felleman, D.J. & Kaas, J.H. (1984). Receptive-field properties of neurons in middle temporal visual area (MT) of owl monkeys. *Journal of Neurophysiology,52*, 488–513.

- Friston, K.J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M.D., & Turner, R. (1998). Event-Related fMRI: Characterizing Differential Responses. *Neuroimage*, *7*, 30-40.

- Huk, A. C., Dougherty, R. F., & Heeger, D. J. (2002). Retinotopy and functional subdivision of human areas MT and MST. *Journal of Neuroscience*, *22*(16), 7195-7205.

- Joo, S. J., Czuba, T. B., Cormack, L. K., & Huk, A. C. (2016). Separate perceptual and neural processing of velocity-and disparity-based 3D motion signals. *Journal of Neuroscience*, *36*(42), 10791-10802.

- Kaestner, M., Maloney, R.T., Wailes-Newson, K., Bloj, M., Harris, J.M., Morland, A.B., Wade, A.R. (*in press*). Distinct response patterns for achromatic and S-cone inputs to cortical motion in depth mechanisms. *Proceedings of the National Academy of Sciences of the United States of America.*

- Kay, K.N., Winawer, J., Mezer, A., & Wandell, B.A. (2013). Compressive spatial summation in human visual cortex. *Journal of Neurophysiology, 110*(2), 481-494.

- Kolster, H., Janssens, T., Orban, G. A., & Vanduffel, W. (2014). The retinotopic organization of macaque occipitotemporal cortex anterior to V4 and caudoventral to the middle temporal (MT) cluster. *Journal of Neuroscience*, *34*(31), 10168-10191.

- Kolster, H., Peeters, R., & Orban, G. A. (2010). The retinotopic organization of the human middle temporal area MT/V5 and its cortical neighbours. *Journal of Neuroscience*, *30*(29), 9801-9820.

- Kolster, H., Mandeville, J. B., Arsenault, J. T., Ekstrom, L. B., Wald, L. L., & Vanduffel, W. (2009). Visual field map clusters in macaque extrastriate visual cortex. *Journal of Neuroscience*, *29*(21), 7031-7039.

- Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., & Baker, C. I. (2009). Circular analysis in systems neuroscience – the dangers of double dipping. *Nature Neuroscience, 12*(5), 535–540.

- Leite, F.P., Tsao, D., Vanduffel, W., Fize, D., Sasaki, Y., Wald, L.L., Dale, A.M., Kwong, K.K., Orban, G.A., Rosen, B.R., Tootell, R.B.H., Mandeville, J.B. (2002). Repeated fMRI using iron oxide contrast agent in awake, behaving macaques at 3Tesla. *Neuroimage, 16*(2), 283– 294

- Likova, L. T., & Tyler, C. W. (2007). Stereomotion processing in the human occipital cortex. *Neuroimage*, *38*(2), 293-305.

- Logothetis, N.K. & Wandell, B.A. (2004). Interpreting the BOLD signal. *Annual Review of Physiology, 66*, 735-769.

- Maunsell, J. H., & Newsome, W. T. (1987). Visual processing in monkey extrastriate cortex. *Annual review of neuroscience*, *10*(1), 363-401.

- Maunsell JH, Van Essen DC (1983) Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. J Neurophysiol 49:1148–1167.

- McLaren, D.G., Kosmatka, K.J., Kastman, E.K., Bendlin, B.B., & Johnson, S.C. (2010). Rhesus macaque brain morphometry: A methodological comparison of voxel-wise approaches. *Methods, 50*, 157–165.

- McLaren, D.G., Kosmatka, K.J., Oakes, T.R., Kroenke, C.D., Kohama, S.G., Matochik, J.A., Ingram, D.K., Johnson, S.C. (2009). A population-average MRI-based atlas collection of the rhesus macaque. *Neuroimage, 45*, 52–59.

- Nelissen, K., Vanduffel, W., Orban, G.A. (2006) Charting the Lower Superior Temporal Region, a New Motion-Sensitive Region in Monkey Superior Temporal Sulcus. *Journal of Neuroscience, 26*, 5929– 5947.

- Newsome, W. T., & Paré, E. B. (1988). A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *Journal of Neuroscience*, *8*(6), 2201-2211.

- Orban, G. A. (2002). Functional MRI in the awake monkey: The missing link. *Journal of Cognitive Neuroscience*, *14*(6), 965-969.

- Orban, G. A., Fize, D., Peuskens, H., Denys, K., Nelissen, K., Sunaert, S., ... & Vanduffel, W. (2003). Similarities and differences in motion processing between the human and macaque brain: evidence from fMRI. *Neuropsychologia*, *41*(13), 1757-1768.

- Rokers, B., Cormack, L. K., & Huk, A. C. (2009). Disparity-and velocity-based signals for three-dimensional motion perception in human MT+. *Nature neuroscience*, *12*(8), 1050.

- Sakata, H., Taira, M., Kusunoki, M., Murata, A., Tanaka, Y. (1997). The TINS lecture; The parietal association cortex in depth perception and visual control of hand action. *Trends in Neuroscience, 20,* 350–357.

- Sanada, T. M., & DeAngelis, G. C. (2014). Neural representation of motion-in-depth in area MT. *Journal of Neuroscience*, *34*(47), 15508-15521.

- Shikata, E., McNamara, A., Sprenger, A., Hamzei, F., Glauche, V., Büchel, C., & Binkofski, F. (2007). Localization of human intraparietal areas AIP, CIP, and LIP using surface orientation and saccadic eye movement tasks. *Human Brain Mapping, 29*(4), 411-421.

- Takeuchi, H., Taki, Y., Hashizume, H., Sassa, Y., Nagase, T., Nouchi, R. & Kawashima, R. (2011). Effects of training processing speed on neural systems. *Journal of Neuroscience, 31*(34), 12139-12148.

- Van Dromme, I. C., Premereur, E., Verhoef, B. E., Vanduffel, W., & Janssen, P. (2016). Posterior parietal cortex drives inferotemporal activations during three-dimensional object vision. *PLoS biology*, *14*(4), e1002445.

- Vanduffel, W., Fize, D., Mandeville, J. B., Nelissen, K., Van Hecke, P., Rosen, B. R., ... & Orban, G. A. (2001). Visual motion processing investigated using contrast agent-enhanced fMRI in awake behaving monkeys. *Neuron*, *32*(4), 565-577.

- Vanduffel, W., Fize, D., Peuskens, H., Denys, K., Sunaert, S., Todd, J. T., & Orban, G. A. (2002). Extracting 3D from motion: differences in human and monkey intraparietal cortex. *Science*, *298*(5592), 413-415.

- Van Essen, D.C., Drury, H.A. Dickson, J., Harwell, J., Hanlon, D., Anderson, C.H. (2001). An integrated software suite for surface-based analyses of cerebral cortex. *Journal of the American Medical Informatics Association, 8*(5), 443– 459.

- Wall, M.B. & Smith, A.T. (2008). The representation of egomotion in the human brain. *Current Biology, 18*(3), 191-194.