



Published in final edited form as:

*Neuroimage*. 2015 April 1; 109: 84–94. doi:10.1016/j.neuroimage.2015.01.012.

## Functional MRI mapping of dynamic visual features during natural viewing in the macaque

Brian E. Russ<sup>1</sup> and David A. Leopold<sup>1,2</sup>

<sup>1</sup>Section on Cognitive Neurophysiology and Imaging, Laboratory of Neuropsychology, National Institute of Mental Health, National Institutes of Health, Bethesda, MD 20892

<sup>2</sup>Neurophysiology Imaging Facility, National Institute of Mental Health, National Institute of Neurological Disorders and Stroke, National Eye Institute, National Institutes of Health, Bethesda, MD 20892

### Abstract

The ventral visual pathway of the primate brain is specialized to respond to stimuli in certain categories, such as the well-studied face selective patches in the macaque inferotemporal cortex. To what extent does response selectivity determined using brief presentations of isolated stimuli predict activity during the free viewing of a natural, dynamic scene, where features are superimposed in space and time? To approach this question, we obtained fMRI activity from the brains of three macaques viewing extended video clips containing a range of social and nonsocial content and compared the fMRI time courses to a family of feature models derived from the movie content. Starting with more than two dozen feature models extracted from each movie, we created functional maps based on features whose time courses were nearly orthogonal, focusing primarily on faces, motion content, and contrast level. Activity mapping using the face feature model readily yielded functional regions closely resembling face patches obtained using a block design in the same animals. Overall, the motion feature model dominated responses in nearly all visually driven areas, including the face patches as well as ventral visual areas V4, TEO, and TE. Control experiments presenting dynamic movies, whose content was free of animals, demonstrated that biological movement critically contributed to the predominance of motion in fMRI responses. These results highlight the value of natural viewing paradigms for studying the brain's functional organization and also underscore the paramount contribution of magnocellular input to the ventral visual pathway during natural vision.

### Keywords

Natural Vision; Face Patches; Visual Cortex; Macaque

---

Corresponding Author: Brian E. Russ, National Institute of Mental Health, 49 Convent Dr. Rm. 1B80, MSC 4400, Bethesda, Maryland 20892, brian.russ@nih.gov.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

## INTRODUCTION

During an animal's natural visual experience, the brain receives information about the environment in a manner that differs from that in most electrophysiological or fMRI experiments. For example, when a macaque monkey observes its conspecifics, visual input consists not of a series of isolated presentations, but of an evolving thread of dynamic, high- and low-level features superimposed in space and time. While it is straightforward to record neural or fMRI activity while monkeys observe natural movies, few such experiments have been carried out, likely because of the challenges inherent in the data analysis and interpretation. However, given that the brain evolved and operates under such conditions, it may be of great value to use natural paradigms that can complement more conventional approaches to assess functional responses in the brain.

Human fMRI experiments have begun to develop and apply such paradigms, often by having subjects watch commercial movies. Various experiments have used voxel time courses to assess the shared signal variation across brain areas (Bartels and Zeki, 2005; 2004), map intersubject correlations (Hasson et al., 2004; 2008), and measure the degree to which brain activity is predicted by certain features (Bartels et al., 2008; Hanson et al., 2007; Huth et al., 2012). These studies have shown that, despite the challenges of superimposed features, it is possible to create functional maps and assess aspects of functional brain organization under natural viewing conditions. One study has gone further, matching the voxel time courses in the human brain to those in macaques watching the same movies with the aim of establishing homological correspondence between the species (Mantini et al., 2013; 2012).

The macaque extrastriate visual cortex is characterized by having a large number of specialized areas. Visual signals are analyzed by regions apparently dedicated for the processing of motion (Maunsell and Van Essen, 1983), complex spatial structure (Fujita et al., 1992), and spatial cues (Andersen, 1985). Testing with fMRI has identified functional networks, such as that specialized for the processing of faces (Tsao et al., 2003). Does this apparent division of labor, derived from conventional testing, govern neural activity in the macaque's brain during natural viewing?

Here we explore this question by measuring fMRI activity in macaques freely viewing extended natural videos containing diverse social and nonsocial content. Using a family of feature models extracted from each video, we assessed the relative contribution of different visual features throughout the brain. We first show that face feature models yield maps that bear striking similarity to the face patches identified using conventional block design. We next show that in the face patches, as well as in neighboring regions of V4, TEO and TE, the motion feature models, whose time courses are largely uncorrelated with the face feature models, are the primary drivers of the fMRI voxel time courses. These findings demonstrate that natural viewing paradigms can be valuable assays of functional specialization in the macaque brain, and additionally underscore the strong contribution of magnocellular input to ventral stream visual processing during natural vision.

## METHODS

### Subjects

Three adult female rhesus monkeys (M1, M2, M3) participated in the study. Prior to training, monkeys were implanted with a custom-designed and fabricated fiberglass headpost, which was used to immobilize the head during testing. All procedures were approved by the Animal Care and Use Committee of the US National Institutes of Health (National Institute of Mental Health) and followed US National Institutes of Health guidelines. Surgery was performed using sterile procedures (see Maier et al (2008) for details on surgical procedures). The analgesics ketoprofen (2 mg/kg twice daily) and ibuprofen (100 mg twice daily) were sequentially administered for three and four days, respectively, following each surgery. Two of the female rhesus (age 5–6 at the time of the study) were pair-housed with other animals and naïve animals prior to participation in the current study. The third animal (age 10 at the time of the study) was singly housed and had participated in fMRI and neurophysiological experiments prior to the current study. During participation in the experiment, the animals were on water restriction and received their daily fluid intake during their daily testing (see below). Each subject's weight and hydration level was monitored closely and maintained throughout the experimental testing phases.

### fMRI Scanning

Functional magnetic resonance images were collected while the monkeys were engaged in the natural viewing task. Subjects participated in up to 12 5-minute trials per scanning session. Structural and functional images were acquired in a 4.7 tesla, 60-cm vertical scanner (Bruker Biospec, Ettlingen, Germany) equipped with a Bruker S380 gradient coil. Animals sat upright in a specially designed chair and viewed the visual stimuli projected onto a screen above their head through a mirror. We collected whole brain images with an 8 channel transmit and receive radiofrequency coil system (Rapid MR International, Columbus, OH). Functional echo planar imaging (EPI) scans were collected as 40 sagittal slices with an in-plane resolution of 1.5 x 1.5 mm and a slice thickness of 1.5 mm. Monocrystalline iron oxide nanoparticles (MION), a T2\* contrast agent, was administered prior to the start of each scanning session. MION doses were determined independently for each subject to attain a consistent drop in the signal intensity of approximately 60% (Leite et al., 2002), which corresponded to ~8–10 mg/kg MION. The repetition time (TR) was 2.4 s and the echo time (TE) was 12 ms. For each 5-minute presentation of a video clip, either 125 or 250 whole brain images were collected, corresponding to 5 and 10 minutes of data collection. During the 10 min scans, the video appeared on the screen after 2.5 min of fMRI acquisition during rest and disappeared at 7.5 min.

### Video Stimuli

The stimuli used in this study consisted of eighteen 5-minute videos. Fifteen videos depicted conspecifics and heterospecifics engaged in a range of activities in natural settings. Scenes were taken from a set of commercially produced nature documentaries (see Video S1 for example stimulus). The remaining three videos, which we termed as the non-biological motion stimuli, had dynamic scenes without a single animal present. These included footage of natural events such as tornados, volcano eruptions, and avalanches. In a typical session,

subjects viewed 3–4 different movies each repeated 3–4 times and presented in pseudorandom order (Figure 1A).

### Training and Experimental Design

Prior to the experimental sessions, the subjects were trained on a fixation task, in which they were required to fixate a  $0.2^{\circ}$ – $0.5^{\circ}$  white dot and hold this fixation for 2 s to receive a juice reward. Viewing of the videos was preceded by an eye movement calibration task, where the animal directed its eyes to a spot appearing at different points on the screen. Such calibration was repeated after every second five-minute trial to reaffirm fixation accuracy throughout the session. Eye position was recorded using an MR-compatible infrared camera (MRC Systems, Heidelberg, Germany) fed into an eye tracking system (SensoMotoric Instruments GmbH, Teltow, Germany). Horizontal and vertical eye position was sampled and saved at 200 Hz.

Each natural viewing trial began with the brief 500 ms presentation of a white central point surrounded by a 10-degree diameter annulus, with the extent of the annulus indicating the spatial region within which the animal could earn reward. Since the annulus was approximately as large as the movie itself, the animals were free to view the content of the video. As long as their eyes remained directed to the movie region, they received a drop of juice reward every two seconds. Pilot studies revealed that while the subjects would watch new videos without any reward incentive, the reward ensured that they would continue to view the same video even after several presentations.

Subject M1 participated in 39 sessions of the repeated natural viewing task, in 14 of which the movie was flanked by 2.5 min rest periods. Subject M1 watched each of the 15 movies an average of ~20 times over the course of the study. Subject M2 participated in 27 sessions total, 11 of which had the 2.5 min flanking periods. Subject M2 watched each movie an average of ~15 times over the study period. Subject M3 participated in 3 sessions of the experiment, during which she viewed 9 of the movies 3 times each.

### Face Localizer Block Design

Face patches were localized and defined using a block design contrasting monkey faces (Gothard et al., 2004), and phase scrambled version of those faces. Each run consisted of eight alternating 48 s blocks. Each image subtended  $12^{\circ}$  visual angle, with a  $0.4^{\circ}$  red fixation point in the center, and was presented for 2 s. Subjects were free to scan the images, and were rewarded every 2 s as long as their gaze was maintained within a  $6^{\circ}$  radius centered fixation window. Over the course of data collection, each subject participated in multiple face localizer runs (17 in M1, 37 in M2, and 4 in M3). Functional maps were produced for each monkey by concatenating all runs and computing the response contrast (t-value) between the two stimulus conditions.

### Construction of Feature Models

From each 5-minute movie, we extracted a family of high- and low-level feature models to be used in the main analyses of the paper. Low-level visual features included models related to luminance, contrast, and motion and were computed automatically using algorithms

applied to the video image sequence in MATLAB. Higher-level visual features included models related to animals, such as the presence of faces and body extremities and various behaviors, and were coded by the experimenters on a frame by frame basis at four frames per second.

Based on a preliminary analysis of both the correlation between different feature models and the feature models in creating functional maps, we pruned down an initial array of nearly two dozen feature models to six visual models that were largely uncorrelated in their time courses (Figures 1B & 1C). These consisted of three low-level feature models (*motion*, *luminance* and *contrast*) and three high-level feature models (*faces*, *animals*, and *extremities*), whose maximum level of correlation was  $<0.3$ . Additionally, a behavioral model of saccade frequency was computed for each subject.

High-level features were scored by a human rater, who evaluated the content of individual frames with a sampling of 4 frames per second (total of 1200 time points for each coded feature per movie). A broad range of stimulus features ( $>50$ ) were scored in the initial rating, most of which were either too sparse to be of use in our analysis or strongly correlated with one another. In preliminary analysis we pared the initial coding down to three. The *face* model consisted of a binary function that was one when a single face was present on the screen and zero otherwise. This model was strongly correlated with a number of other face-based models we computed initially, such as measures of the number of faces present. For the *animal* and extremity models, we counted the total number of animals, or extremities (i.e. arms and legs) on the screen in each frame and constructed the models based on the square root transforms of these counts.

Low-level feature models were computed by algorithms applied to the movies automatically using MATLAB. In the initial extraction of features, multiple different algorithms were used to create the *motion* feature model. In pilot analysis, we compared total motion, defined as the mean instantaneous speed in the image plane, and local motion, computed according to an algorithm for determining motion within center-surround receptive fields (Itti, 2005; Itti et al., 1998). As these measures were strongly correlated, we opted to use the total motion for our main analysis. *Luminance* was calculated by measuring the average pixel intensity value for each frame of the movie. *Contrast* was calculated using a Weibull fit of the pixel intensity histogram based on the properties of the early visual system (Ghebreab et al., 2009; Groen et al., 2012; Scholte et al., 2009). This measure of contrast is associated with the average strength of the edges present in a given frame.

One model of saccade behavior was also included in the analysis. To this end, we included an estimation of mean instantaneous saccade frequency over the duration of each movie. To compute this, we detected the instance of saccades on each trial whose peak velocities exceeded (40 deg/sec). Binning these for each fMRI TR, and then dividing by the bin width and total number of trials provided an estimation of the time varying mean saccade frequency associated for each movie. Note that this analysis focuses on the average saccade frequency and does not pertain to trial-unique patterns of eye movements, which is the topic of a separate report. The overall mean saccade rate across movies was 1.7 sac/sec for M1 and 1.4 sac/sec for M2.

All feature models were down-sampled using the *decimate* function in MATLAB to match the same temporal resolution as the fMRI data (i.e. 1 sample per 2.4 seconds), convolved with an estimation of the hemodynamic MION response function, and concatenated across all movies for comparison with the concatenated average fMRI data.

### fMRI Analysis

All fMRI data were analyzed using custom-written MATLAB (MathWorks, Natick, MA) programs as well as the AFNI/SUMA software package developed at NIH (Cox, 1996). Raw images were first converted from Bruker into AFNI data file format. Motion correction algorithms were applied to each EPI time course using the AFNI function *3dvolreg*, followed by correction for static magnetic field inhomogeneities using the PLACE algorithm (Xiang and Ye, 2007). Each session was then registered to a template session, allowing for the combination of data across multiple testing days. The first 7 TRs (16.8 s) of each movie were not considered in the analysis so as to eliminate the hemodynamic onset response associated with the initial presentation of each video.

The inherent temporal flow of an extended video clip poses challenges for using conventional fMRI design to study the neural mechanisms of natural vision. Normally, functional maps are created based on *a priori* models derived from the stimulus presentation sequence, for example in block and event-related designs, where the time courses of all responsive voxels are usually assumed to follow the on/off temporal dynamics of stimulus presentation. Since extended video stimuli add their own inherent temporal structure to the responses of individual voxels, which will differ throughout the brain based on regional response selectivity, alternative analysis strategies are required. Previous studies have addressed this challenge by developing purely data-driven methods, for example computing the statistical covariation among voxel time courses during movie viewing (Bartels et al., 2008; Bartels and Zeki, 2004), or the consistency in the time courses of corresponding voxels across subjects (Hasson et al., 2008; 2004). Here we applied a hybrid method, extracting feature models from movies, thus following the temporal dynamics of a naturally evolving scene. We applied three main analyses, investigating the correlation with the feature models, using multiple models in a stepwise regression, and investigating the relative contribution of the different models in a region of interest analysis. Each of these is described in the following paragraphs.

### Correlation with each feature model

We compared the time course in the brain to each of the six extracted feature models using Pearson's correlation. The time courses for each voxel were averaged over all viewings and concatenated across the different movies. This analysis led to a family of activity maps (r-values) that provided a portrait of the functional activation associated with each feature. The significance of r-values was assessed by calculating the two-tailed p-value of each feature model's correlation with a voxel using the number of time-points across all movies minus two as the number of degrees of freedom ( $df = 1768$ ). A Bonferroni correction was applied to account for multiple comparisons, yielding an estimate of significance ( $p < 0.05$ , corrected for the total number of voxels in the field of view, which exceeded 90,000) that corresponds to a Pearson's correlation of  $r = \pm 0.1181$ . Maps of r-values were displayed in

either unthresholded form (Figure 3) or with their transparency modulated by the r-value magnitude for purposes of visualization (Figure 2).

### Stepwise regression using multiple feature models

We next compared the relative contribution of different features to the voxel time courses throughout the brain. To do this we computed a stepwise regression for the seven selected feature models. Stepwise regression provides the ability to directly compare the amount of contribution of each feature, while ensuring that the amount of variance explained is not being over fit. We chose to use stepwise regression, which is a modified version of the commonly used general linear model (GLM) because this method from the outset identified the most relevant models from a large initial number of candidates. Narrowing down the candidates using GLM would yield a similar result, but would require further steps, since the GLM fitting provides a beta coefficient for all models, and does not take redundancy into account. The stepwise regression consisted of an iterative regression sequence. In each cycle, all remaining feature models (starting with the seven selected models) were independently regressed against each time course. With each cycle, a new feature model was added to the overall model for a given voxel if its regression explained the highest percent of the signal variance *and* this value was above 1%. This method provides an inherent threshold for data visualization, where voxels for which a given feature model did not explain 1% of the variance are without any value. Note that the seven feature models used in the main analysis were from the outset selected based on a criterion of minimizing inherent correlation (see Figure 1C). In addition, our application of stepwise regression guarded against the redundant contribution of different feature models.

### Region of Interest Analysis

The stepwise regression analysis afforded the opportunity to compare the relative contribution of the different stimulus features within voxels throughout the brain, serving as a complementary analysis to the maps in Figure 3. We approached this comparison by creating regions of interest (ROI) across the striate and extrastriate visual cortex. This was achieved in two ways. First, functional ROIs were created for the individual face patches by selecting all voxels that were above a threshold of  $t = 5$  based on the results from the block design. In the prefrontal cortex this threshold was lowered to  $t = 2$ , as the signals were inherently weaker. Next, structural ROIs for eight visual regions were selected based on the Saleem and Logothetis monkey atlas (Saleem and Logothetis, 2012). To compare across features, we selected 10 voxels from each ROI with the highest total variance explained from the stepwise model and computed the average value for each of the seven features. Post-hoc two-tailed t-tests were then performed to investigate whether there was a significant difference between the amount of variance explained by the motion and face features.

## RESULTS

The main goal of the study was to measure the contribution of different stimulus features to fMRI responses throughout the brain during natural viewing. The fMRI data were obtained from three monkeys, who repeatedly viewed up to eighteen five-minute natural videos inside

the bore of a 4.7T vertical scanner. The movies varied in their content, but generally depicted conspecific and heterospecific monkeys, humans, and other animals engaged in a variety of social and nonsocial behaviors (see Video S1 for a depiction of the movie content). The subjects viewed individual movies between 3 and 40 times (mean = 14), with fMRI data collected during a total of 728 viewings over a period of several months (389 for M1, 309 for M2, 30 for M3). In a typical session, subjects viewed 3–4 different movies each repeated 3–4 times and presented in pseudorandom order (Figure 1A). Based on tracking their eye movements, it was evident that the monkeys readily explored the movies, even during repeat presentations of the same movie, consistent with previous reports (Shepherd et al., 2010). Across all viewings, gaze was directed toward the screen on average 86% of the time (M1: 87%, M2: 87%, M3: 84%). Since gaze was directed toward the movies most of the time, all trials were included in the mean time courses for each movie presented below. The link between eye movements and fMRI responses across trials is considered briefly here and is investigated in detail in a parallel study. We optimized the estimation of each voxel's representative movie-driven time course by computing the mean response across all presentations of each movie. Voxels across the brain differed markedly in their time courses in response to the movies, making it possible to assess the relative contribution of particular stimulus features to the responses in each area during natural viewing.

Analysis of fMRI responses, including the creation of functional maps, involved comparing the individual voxels' mean time courses to a family of feature models extracted from the movies. Based on pilot analysis, our study focused on six main models (Figure 1B) selected from a much larger initial set of candidates. A critical aspect of the selected feature models is that their time courses were largely uncorrelated with one another (Figure 1C, see **Methods** for more details on the generation and selection feature models). We begin by focusing on one feature, faces, and show that the evolving face content of the movie can be used to map the macaque face patches that closely resemble those obtained with a conventional block design. We next broaden the exploration to the six nearly orthogonal visual feature time courses, demonstrating that the very same fMRI data sets offer a range of distinct and robust activity maps reflecting the processing of faces, contrast, and motion content. Finally, using a quantitative comparison of the variance explained by each feature, we demonstrate that stimulus motion dominates responses across the extended visual system during natural viewing, including in the face patches and ventral object pathway.

### Mapping face patches from natural videos

As an initial test of the feature mapping approach, we first asked whether the well-described macaque face patches (Tsao et al., 2003; Ungerleider and Bell, 2011) could be mapped by comparing voxel time courses across the brain to the time course of the extracted face feature model. Previous human studies have found that voxels in face-selective regions show their highest responses to faces appearing in a movie (Hasson et al., 2004), suggesting that such a mapping approach might be possible.

Indeed, computing the brain-wide correlation between fMRI activity and the face feature model revealed discrete patches of activity in the lower bank of the STS and the prefrontal cortex (Figure 2B & 2D). The locations of these patches closely resembled those reported in



previous macaque fMRI studies (Tsao et al., 2003; 2008b; Ungerleider and Bell, 2011), and numbered 36 face-selective patches in six hemispheres of three monkeys (see Table 1). To more closely evaluate the relationship between these activity patterns and the previously described face patches, we also performed a conventional block design in the same animals (Figure 2A & 2C). Comparing these two modes of data collection revealed striking overlap in nearly all identified face patches in the three monkeys. The face feature model used here was based on the appearance of isolated faces on the screen: pilot experiments determined that other related face feature models yielded similar results (see **Methods**). In addition, the natural viewing approach led to the identification of a few areas outside the face patches (see Table 1). While the maps in Figure 2 were computed using a large number of trials, we found that as little as 15 minutes of movie viewing led to the reliable mapping of activity in the face patches (Supplemental Figure 1).

This mapping of the face patches under natural viewing demonstrates that it is possible to use this approach to gauge aspects of the brain's functional organization despite multiple stimulus features being superimposed in space and time. The close correspondence to face patches further suggests that data collected during conventional fMRI paradigms provides a reasonable path toward understanding how the brain responds under more natural viewing conditions. It should be noted, however, that while this method was able to reveal the position of the face patches, it yielded correlation that seldom exceeded a value of 0.3 in the face patches, which corresponds to less than 10% of the overall signal variance. We will return to this point later.

### Parallel functional feature maps obtained during natural stimulation

In light of the successful mapping using the face feature model, we next examined the family of maps derived from the same fMRI data using the six different visual feature models. The patterns of correlation are shown on the lateral surface of the brain in Figure 3 (medial surfaces are shown in Supplemental Figure 2) and illustrate the spatially distinct contribution of the different features in one hemisphere of subject M2 (Subject M1 shown in Supplemental Figure 3). The upper left map corresponding to the face feature model illustrates the circumscribed nature of the face patches extracted using this method. It shows, in addition, several regions of negative correlation distributed in both the dorsal and ventral visual pathways.

The time course of the selected high-level features other than faces exhibited relatively low correlation with fMRI activity during movie viewing. While both the *extremities* and *animals* feature models yielded marginally significant correlation in some cortical regions (two-tailed t-test(1768),  $p < 0.05$ , corrected), the overall correlation was much lower than for faces and near zero throughout most of the brain. The exceptions were a few regions of negative correlation observed for the *extremities* feature model, which partially overlapped the negative regions observed with faces.

Mapping the time course of low-level features from the same data set led to notably stronger patterns of correlation across the brain, and including in the ventral visual stream. The *contrast* and *luminance* features were both correlated with responses in the early retinotopic cortical areas V1 and V2. In addition, *contrast* was strongly correlated with fMRI time

courses in ventral V3, ventral V4, TEO and regions of TE in the lateral occipitotemporal pathway. The correlation of voxel time courses with fluctuations in *luminance* and *contrast* was weak or absent elsewhere in the brain, including along the entire STS and prefrontal cortex. Activity changes related to image luminance were largely restricted to the foveal representation of V1, and nearly absent at later stages.

The visual feature that was the most strongly correlated with fMRI fluctuations in response to the movie was *motion* (Figure 3, bottom middle). Correlation values were highest in motion-specialized cortical areas MT, MST, and FST in the caudal superior temporal sulcus (STS), consistent with previous work using biological motion (Bartels et al., 2008; Fischer et al., 2012; Nelissen et al., 2006; Orban and Jastorff, 2013). In those areas, peak correlation coefficients exceeded 0.7, indicating a very close match between some individual voxels and the *motion* time course (see Figure 4A). Strong correlations were also found in area V4, the lateral aspect of the IT cortex and rostral portions of the STS, areas commonly associated with object processing in the ventral visual pathway (see Figure 3, and next section). One region in which the correlation was conspicuously absent was the foveal representation of the primary visual cortex (V1), possibly because of the tendency toward stabilization at the center of gaze, or possibly because of a lower proportion of magnocellular input to V1 from the foveal region of the retina (Schein and de Monasterio, 1987).

Together, these results are important for two reasons. First, they demonstrate that it is straightforward to compute a family of distinct feature-specific functional maps from a single fMRI data set collected under experimental conditions that approximate those of natural vision. Second, they show that the dynamic, movement-related aspects of natural vision are dominant in shaping responses throughout the primate brain, even in areas ostensibly specialized for semantic object categories. In the next section, we use stepwise regression to more rigorously analyze the relative contribution of the different stimulus features to the activity of individual voxels, focusing on the ventral visual pathway.

### Assessing the relative contribution of features underscores the dominance of motion

As a next step, we assessed the relative contribution of individual features to fMRI responses throughout the brain using a stepwise regression. We computed the percentage of variance explained by each of the feature models, including a behavioral measure of saccadic eye-movements, to fMRI time courses. Stepwise regression employs an iterative algorithm to determine the best-fitting linear combination of functions, with each iteration adding a new function (i.e. feature model) to the overall model of the data only if a certain established criterion is met. In our case, this criterion was that the new model must add at least 1% to the total variance to qualify for inclusion (see **Methods**).

Applying this analysis to predefined regions of interest throughout the brain underscored the predominance of motion over other stimulus attributes throughout visually responsive cortical areas (Figures 4B & 4C). As a percentage of the total variance explained, the *motion* regressor outranked all the other regressors throughout most dorsal and ventral stream areas, with the most obvious exception being foveal V1, where *contrast* explained the most variance (Figure 4C). In areas MT and FST voxels, the motion feature model contributed as much as half the overall variance, reflecting the previously mentioned correlation coefficient

of  $\sim 0.7$ , and was frequently the only model meeting the inclusion criterion for the stepwise regression. Finally, the average time course of saccade frequency was virtually uncorrelated with the fMRI responses within these regions.

In the ventral stream areas V4, TEO, and TEad, which are generally associated with object processing, motion was the primary contributor to the fMRI response. Even in the face patches, the contribution of motion was dominant (Figure 4B). In fact, in 25 of the 36 face patches identified in the three monkeys (6 hemispheres total: 14 face patches in Subject M1, 13 in Subject M2, 9 in Subject M3), motion accounted for a similar or higher proportion of the percent variance explained than faces (posthoc t-test comparing *motion* to *faces*,  $p < 0.05$ ).

In the ventral visual cortex, the contribution of motion took the form of a posterior to anterior gradient, both inside and outside the STS. This can be seen in one dimensional depictions of the functional maps, where the mean % variance explained for each of three variables is shown from posterior to anterior regions, for both the lower bank of the STS and the lateral surface of V4/TEO/IT (Figure 4D). On the lower bank of the STS, the percentage of variance explained by motion was highest near area MT (cyan arrow) and declined gradually toward the temporal pole. On the lateral surface, the maximum motion contribution was in area V4 (orange arrow), where the magnitude was similar to MT. Unlike in the STS, where stimulus contrast played almost no role in the fMRI responses, the lateral surface showed a pronounced dependence on contrast, which also took the form of a posterior to anterior gradient. Note that a subset of the face patches is visible in Figure 4D (black arrows), and the relative contribution of face and motion for these anterior-posterior positions can be estimated.

To investigate the sensitivity of these results to the specific models selected for the stepwise regression, we applied the same approach to different combinations features. We found that the activity maps pertaining to motion, faces, and contrast were remarkably robust. Even when twenty-one feature models were included into the stepwise regression, these three features accounted for the highest proportion of the overall variance, led to the same spatial pattern across the brain, and contributed in the same way to individual voxels (Supplemental Figure 4). This finding demonstrates that our findings reflect a robust aspect of the brain's responses during natural vision and are not strongly dependent on to the specific choices made in extracting stimulus features.

### Biological versus nonbiological motion

The correlation and stepwise regression results both suggest that movement is critical in shaping fMRI responses during natural vision. To explore the role of biological versus non-biological movement in the motion-based responses, we applied the same analysis to data collected when monkeys watched natural movies consisting of motion that was not generated by biological movements. We compared the results of the regression analysis for these non-biological movies to the same amount of data collected during the viewing of the social movies used for the main portion of the study. The new movies consisted of scenes of disasters such as tornados, volcano eruptions, avalanches, but no animals. Subjects M1 and M2 watched each of the non-biological motion movies 12 and 14 times respectively. While the subjects may have been somewhat less enthusiastic about watching these movies than

the social movies, they did so in exchange for juice reward, which we verified based on their gaze position (mean percent time eyes were directed at the movie equaled 83.6% (M1 = 83.7%; M2 = 83.6%).

The analysis revealed that movie motion caused by non-biological events was largely ineffective in stimulating the visual pathways (Figure 5), with the most pronounced differences between the two movie types observed in the STS and in area V4. While the percentage of variance explained in these regions during the social videos often exceeded 30%, that obtained during the non-social movies seldom exceeded 10%. Even in the motion specialized areas MT and FST, the *motion* feature regressor contributed much less to the response time courses with the non-social movies. While *contrast* maps were highly similar between the two types of movies, *motion* only contributed strongly to the ventral visual pathway for movies that contained biological movement.

## DISCUSSION

We used fMRI data collected during the free viewing of natural movies to investigate how the brain processes visual features that co-occur during natural vision. We found that we could readily map the locations of the face patches across three animals using a model based on the presence of faces. Additionally, we found that motion dominated the responses throughout the ventral visual stream, both within known motion selective regions in the lower bank of the STS (Bartels et al., 2008; Fischer et al., 2012; Jastorff et al., 2012; Nelissen et al., 2006; Orban and Jastorff, 2013) and for traditionally object selective cortex along the lateral bank of IT, including in the face patches. In the following sections we discuss the particular importance of movement signals for social vision, the contribution of magnocellular input to the ventral visual pathway, and the trade-offs associated with studying brain function using natural paradigms.

### Movement cues in social visual perception

Previous experiments have shown that the ventral visual pathway is critical for the high-level interpretation of visual structure. These neural populations are typically studied by measuring their responses to briefly flashed static images. Single-unit recordings have found that cells in the inferior temporal (IT) cortex, are selectively activated by complex patterns, with many apparently specialized for socially relevant stimuli such as faces and bodies (Desimone et al., 1984; Popivanov et al., 2014; Tsao et al., 2006). fMRI experiments have mapped the spatial organization of selective responses in the ventral pathway, specifying a network of face-selective “patches” in the macaque temporal cortex (Bell et al., 2009; Tsao et al., 2006). Other stimulus categories, such as bodies, and biological motion are also known to evoke clustered single-unit and fMRI responses (Bell et al., 2009; Jastorff et al., 2012; Oram and Perrett, 1996). Such basic selectivity is upheld when animals direct their gaze to stimuli embedded in cluttered scenes (DiCarlo and Maunsell, 2000; Sheinberg and Logothetis, 2001). Additionally, movement of a stimulus often increases its fMRI responses in the ventral pathway (Bartels et al., 2008; Nelissen et al., 2006; Orban and Jastorff, 2013). For example, in certain face patches, dynamic faces elicit stronger fMRI responses than static faces (Furl et al., 2012; Polosecki et al., 2013). During natural vision, these and other

visual features are superposed within complex and dynamic scene. It has thus long been clear that movement is an important visual feature in the ventral visual pathway. Nonetheless, our findings that motion contributes more than any other feature to ventral stream responses, and that this dominance only pertains to videos containing social content, draws new attention to the ventral visual pathway's profound entrainment to the movements of dynamic natural visual stimuli.

From the perspective of a visual system, animals are unusual stimuli. Their bodies and faces frequently undergo spontaneous and non-rigid transformations, which are often of great significance to an observer, as they provide immediate and direct cues about others' actions and allow for some element of behavioral prediction. For animals with high visual acuity, such as primates, movement patterns are an important cue for recognizing the identity or actions of an individual at a distance. Given the importance of bodily movement, it is likely there has been consistent evolutionary pressure on the visual brain to specialize in the decoding of actions. In the case of primates, including humans, the brain is exquisitely sensitive to subtle movements, reflecting the primary role of vision in social perception (Leopold and Rhodes, 2010; Parr, 2011).

In our data, most face patches showed either stronger or equivalent responses to motion content than to face content. Since the time courses of the face and motion feature models were uncorrelated ( $r = -0.0275$ ,  $p > .05$ ), one must conclude that the fMRI signal in these regions were more strongly shaped by motion than by the simple presence of faces. This conclusion extends findings from two recent studies in the macaque, which compared fMRI responses to static and dynamic faces (Furl et al., 2012; Polosecki et al., 2013). Nonetheless, it is important to point out that this motion sensitivity was not restricted to face patches, but was expressed as a posterior to anterior negative gradient, both on the lower bank of the STS and on the lateral surface of IT cortex. While the basis of this gradient is not known, it could reflect a transition from more structural to a more semantic representation of complex biological stimuli, with motion cues used primarily to derive structural information.

Within certain faces patches, particularly the more posterior ones (PL and ML), the extremities feature sometimes accounted for as much variance as the face feature. The extremities feature also accounted for a modest but significant portion of the variance in other visual areas. These responses are likely to reflect biological movement associated with the limbs, which is known to be an effective stimulus for many cortical areas (Bartels et al., 2008; Fischer et al., 2012; Jastorff et al., 2012; Nelissen et al., 2006; Orban and Jastorff, 2013), including body areas bordering the face patches of the STS (Bell et al., 2009; Popivanov et al., 2014; Tsao et al., 2003; 2008a). In the current study, we did not find any regions that were strongly associated with the presence of extremities in the movie (see Figure 3). This result could stem from an inherent lack of regional selectivity for the extremity features we chose, including hands, feet, and limbs. While individual neurons may have selectivity for these features, it may be organized across the cortical surface at a spatial scale that makes it impossible to identify an "extremity area" (Popivanov et al., 2014). Future analysis may benefit from considering the contents of the fovea based on eye tracking, for example, considering whether the animal is looking at a hand or limb at each moment, as opposed to simply considering whether one is present on the screen.

## Magnocellular contribution to the “what” pathway

Motion’s contribution to neural responses in the ventral stream was recognized early, including in the first recordings from the IT cortex (Gross et al., 1979; 1972). Subsequent single unit studies discovered that neurons in certain regions of the STS appeared specialized for biological movements (Perrett et al., 1985). Such movements included active head rotation, limb movement, locomotion, as well as movable features of the face, such as the direction of eye gaze. More recent studies have emphasized the integration of static form and dynamic movement in neurons in the STS visual areas (Beauchamp et al., 2003; Engell and Haxby, 2007; Jastorff et al., 2012; Oram and Perrett, 1996; Puce and Perrett, 2003; Singer and Sheinberg, 2010). While the importance of motion for responses in the STS is thus expected (Bartels et al., 2008; Bartels and Zeki, 2004; Fischer et al., 2012; Jastorff et al., 2012; Nelissen et al., 2006; Orban and Jastorff, 2013), the domination of motion in fMRI responses throughout the entire ventral stream, including area V4 and lateral aspects of TEO and TE that are strongly associated with object vision, could not have been predicted based on previous work.

By what neural pathways do these lateral form and object-selective regions receive motion signals? While there are a number of routes carrying visual information from the retina to the ventral visual stream (Kravitz et al., 2013), movement-related signals are ultimately thought to draw heavily upon the magnocellular pathway (Nassi and Callaway, 2009). Disruption of the magnocellular LGN layers severely disrupts the perception of moving or flickering stimuli (Schiller et al., 1990), and strongly reduces activity in area MT (Maunsell et al., 1990). One possibility is that the observed motion-sensitive signals in the ventral stream pass through MT. This prospect is supported by anatomical studies showing that MT sends projections rostrally within the STS including to area FST (Ungerleider and Desimone, 1986), which sends prominent projections to regions in the fundus, lower bank of the STS (Boussaoud et al., 1990), and the lateral surface of IT (Distler et al., 1993). This MT pathway has been used to explain the persistence of visual responses with complex selectivity in IT following the ablation of areas V4 and TEO (Bertini et al., 2004). Another possibility is that motion signals reach V4 more directly, without passing through MT. Area V4 receives roughly equal shares of input from magnocellular and parvocellular channels, as revealed by combined recordings and inactivation studies (Ferrera et al., 1994), and projects widely to ventral stream areas (Kravitz et al., 2013). Thus there are multiple parallel routes by which magnocellular signals can shape responses in the so-called “what” pathway.

## Trade-offs inherent in natural viewing paradigms

In designing experiments to probe brain functions, researchers are forced to navigate a continuum of complexity that taxes the level of experimenter control. At one end of the spectrum, one can present simple, well-characterized stimuli repeatedly to a precise position on the retina, in anesthetized animals. Such controlled experiments have provided the foundation for understanding visual response selectivity in the brain. At the other end of the spectrum, one can imagine an experiment involving free-ranging animals, implanted with large microelectrode arrays, interacting naturally as their neural activity patterns are telemetrically recorded. This mode of data collection would yield brain activity patterns that, while rich and natural in their content, are practically impossible to decipher.

Our study sits at an intermediate position along this continuum, similar to a number of other recent approaches (Bartels and Zeki, 2004; Hasson et al., 2012; Huth et al., 2012). The repeated viewing of natural videos departs from a strictly reductionistic approach but retains sufficient experimental structure to assess and map consistent patterns of brain activity. The method we present in the current paper applies only one approach to brain mapping during natural viewing, using feature models extracted from the movie itself. As such, the resulting brain activity maps are tightly linked to the definition and extraction of the features. Our focus in this study, which involved pruning down an initial set of nearly two dozen models into seven whose variations were well-behaved across the movies and nearly orthogonal in their time courses, allowed us to assess the relative contribution of a few critical visual features. Importantly, our demonstration that the inclusion of all twenty-one models into the stepwise regression led to the same basic result shows that the result is robust: functional maps of faces and contrast emerged regardless of our initial inclusion criteria, with a preservation of the relative contribution to individual voxels.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We would like to thank Tahir Haque for scoring the videos, Charles Zhu and Frank Ye for assistance with fMRI scanning, Katy Smith, George Dold and David Ide for excellent technical assistance and development, Takaaki Kaneko and Saleem Kadharbatcha for helpful discussion, Hang Joon Jo, Ziad Saad and Colin Reveley for guidance in fMRI analysis, and Chia-chun Hung and Brian Scott for comments on the manuscript. Support for this research was funded by the Intramural Research Program of the National Institute of Mental Health.

## References

- Andersen RA. Oculomotor adaptation: adaptive mechanisms in gaze control. *Science*. 1985; 230:1371–1372. [10.1126/science.230.4732.1371](https://doi.org/10.1126/science.230.4732.1371) [PubMed: 17749678]
- Bartels A, Zeki S. Functional brain mapping during free viewing of natural scenes. *Hum Brain Mapp*. 2004; 21:75–85. [10.1002/hbm.10153](https://doi.org/10.1002/hbm.10153) [PubMed: 14755595]
- Bartels A, Zeki S. Brain dynamics during natural viewing conditions--a new guide for mapping connectivity in vivo. *Neuroimage*. 2005; 24:339–349. [10.1016/j.neuroimage.2004.08.044](https://doi.org/10.1016/j.neuroimage.2004.08.044) [PubMed: 15627577]
- Bartels A, Zeki S, Logothetis NK. Natural vision reveals regional specialization to local motion and to contrast-invariant, global flow in the human brain. *Cereb Cortex*. 2008; 18:705–717. [10.1093/cercor/bhm107](https://doi.org/10.1093/cercor/bhm107) [PubMed: 17615246]
- Beauchamp MS, Lee KE, Haxby JV, Martin A. FMRI responses to video and point-light displays of moving humans and manipulable objects. *Journal of cognitive neuroscience*. 2003; 15:991–1001. [10.1162/089892903770007380](https://doi.org/10.1162/089892903770007380) [PubMed: 14614810]
- Bell AH, Hadj-Bouziane F, Frihauf JB, Tootell RBH, Ungerleider LG. Object representations in the temporal cortex of monkeys and humans as revealed by functional magnetic resonance imaging. *J Neurophysiol*. 2009; 101:688–700. [10.1152/jn.90657.2008](https://doi.org/10.1152/jn.90657.2008) [PubMed: 19052111]
- Bertini G, Buffalo EA, De Weerd P, Desimone R, Ungerleider LG. Visual responses to targets and distracters by inferior temporal neurons after lesions of extrastriate areas V4 and TEO. *Neuroreport*. 2004; 15:1611–1615. [PubMed: 15232293]
- Boussaoud D, Ungerleider LG, Desimone R. Pathways for motion analysis: cortical connections of the medial superior temporal and fundus of the superior temporal visual areas in the macaque. *J Comp Neurol*. 1990; 296:462–495. [10.1002/cne.902960311](https://doi.org/10.1002/cne.902960311) [PubMed: 2358548]

- Cox RW. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res.* 1996; 29:162–173. [PubMed: 8812068]
- Desimone R, Albright TD, Gross CG, Bruce C. Stimulus-selective properties of inferior temporal neurons in the macaque. *J Neurosci.* 1984; 4:2051–2062. [PubMed: 6470767]
- DiCarlo JJ, Maunsell JH. Form representation in monkey inferotemporal cortex is virtually unaltered by free viewing. *Nat Neurosci.* 2000; 3:814–821.10.1038/77722 [PubMed: 10903575]
- Distler C, Boussaoud D, Desimone R, Ungerleider LG. Cortical connections of inferior temporal area TEO in macaque monkeys. *J Comp Neurol.* 1993; 334:125–150.10.1002/cne.903340111 [PubMed: 8408755]
- Engell AD, Haxby JV. Facial expression and gaze-direction in human superior temporal sulcus. *Neuropsychologia.* 2007; 45:3234–3241.10.1016/j.neuropsychologia.2007.06.022 [PubMed: 17707444]
- Ferrera VP, Nealey TA, Maunsell JH. Responses in macaque visual area V4 following inactivation of the parvocellular and magnocellular LGN pathways. *J Neurosci.* 1994; 14:2080–2088. [PubMed: 8158258]
- Fischer E, Bühlhoff HH, Logothetis NK, Bartels A. Visual motion responses in the posterior cingulate sulcus: a comparison to V5/MT and MST. *Cereb Cortex.* 2012; 22:865–876.10.1093/cercor/bhr154 [PubMed: 21709176]
- Fujita I, Tanaka K, Ito M, Cheng K. Columns for visual features of objects in monkey inferotemporal cortex. *Nature.* 1992; 360:343–346.10.1038/360343a0 [PubMed: 1448150]
- Furl N, Hadj-Bouziane F, Liu N, Averbeck BB, Ungerleider LG. Dynamic and static facial expressions decoded from motion-sensitive areas in the macaque monkey. *J Neurosci.* 2012; 32:15952–15962.10.1523/JNEUROSCI.1992-12.2012 [PubMed: 23136433]
- Ghebreab S, Smeulders A, Scholte HS, Lamme V. A biologically plausible model for rapid natural image identification. *Adv Neural Inf Process Syst.* 2009:1–9.
- Gothard KM, Erickson CA, Amaral DG. How do rhesus monkeys (*Macaca mulatta*) scan faces in a visual paired comparison task? *Anim Cogn.* 2004; 7:25–36.10.1007/s10071-003-0179-6 [PubMed: 14745584]
- Groen IIA, Ghebreab S, Lamme VAF, Scholte HS. Spatially pooled contrast responses predict neural and perceptual similarity of naturalistic image categories. *PLoS Comput Biol.* 2012; 8:e1002726.10.1371/journal.pcbi.1002726 [PubMed: 23093921]
- Gross CG, Bender DB, Gerstein GL. Activity of inferior temporal neurons in behaving monkeys. *Neuropsychologia.* 1979; 17:215–229.10.1016/0028-3932(79)90012-5 [PubMed: 111157]
- Gross CG, Rocha-Miranda CE, Bender DB. Visual properties of neurons in inferotemporal cortex of the Macaque. *J Neurophysiol.* 1972; 35:96–111. [PubMed: 4621506]
- Hanson SJ, Hanson C, Halchenko Y, Matsuka T, Zaimi A. Bottom-up and top-down brain functional connectivity underlying comprehension of everyday visual action. *Brain Struct Funct.* 2007; 212:231–244.10.1007/s00429-007-0160-2 [PubMed: 17968590]
- Hasson U, Ghazanfar AA, Galantucci B, Garrod S, Keysers C. Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends Cogn Sci (Regul Ed).* 2012; 16:114–121.10.1016/j.tics.2011.12.007 [PubMed: 22221820]
- Hasson U, Nir Y, Levy I, Fuhrmann G, Malach R. Intersubject synchronization of cortical activity during natural vision. *Science.* 2004; 303:1634–1640.10.1126/science.1089506 [PubMed: 15016991]
- Hasson U, Yang E, Vallines I, Heeger DJ, Rubin N. A Hierarchy of Temporal Receptive Windows in Human Cortex. *Journal of Neuroscience.* 2008; 28:2539–2550.10.1523/JNEUROSCI.5487-07.2008 [PubMed: 18322098]
- Huth AG, Nishimoto S, Vu AT, Gallant JL. A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron.* 2012; 76:1210–1224.10.1016/j.neuron.2012.10.014 [PubMed: 23259955]
- Itti L. Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Visual Cognition.* 2005; 12:1093–1123.10.1080/13506280444000661
- Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on.* 1998; 20:1254–1259.

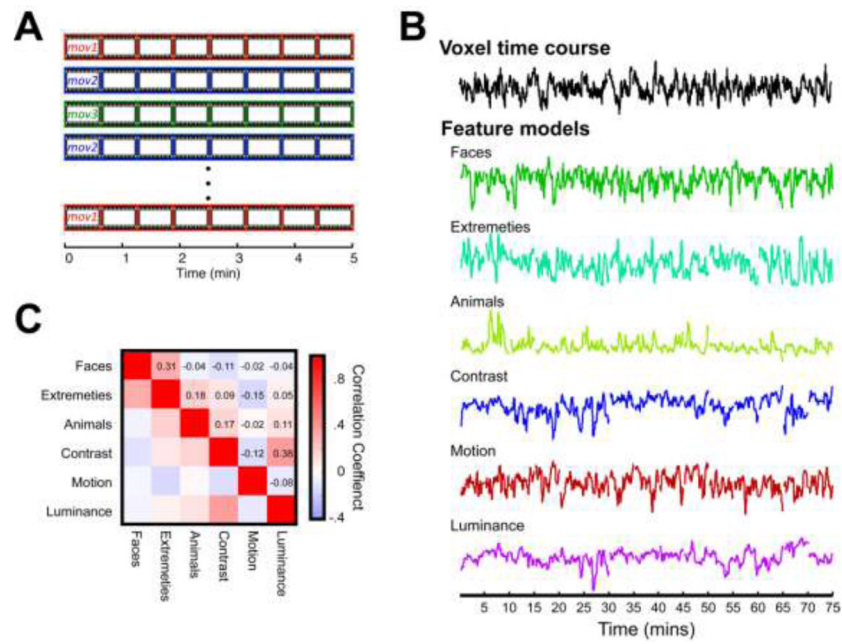


- Jstorff J, Popivanov ID, Vogels R, Vanduffel W, Orban GA. Integration of shape and motion cues in biological motion processing in the monkey STS. *Neuroimage*. 2012; 60:911–921.10.1016/j.neuroimage.2011.12.087 [PubMed: 22245356]
- Kravitz DJ, Saleem KS, Baker CI, Ungerleider LG, Mishkin M. The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends Cogn Sci (Regul Ed)*. 2013; 17:26–49.10.1016/j.tics.2012.10.011 [PubMed: 23265839]
- Leite FP, Tsao D, Vanduffel W, Fize D, Sasaki Y, Wald LL, Dale AM, Kwong KK, Orban GA, Rosen BR, Tootell RBH, Mandeville JB. Repeated fMRI using iron oxide contrast agent in awake, behaving macaques at 3 Tesla. *Neuroimage*. 2002; 16:283–294.10.1006/nimg.2002.1110 [PubMed: 12030817]
- Leopold DA, Rhodes G. A Comparative View of Face Perception. *J Comp Psychol*. 2010; 124:233–251.10.1037/a0019460 [PubMed: 20695655]
- Maier A, Wilke M, Aura C, Zhu C, Ye FQ, Leopold DA. Divergence of fMRI and neural signals in V1 during perceptual suppression in the awake monkey. *Nat Neurosci*. 2008; 11:1193–1200.10.1038/nn.2173 [PubMed: 18711393]
- Mantini D, Corbetta M, Romani GL, Orban GA, Vanduffel W. Evolutionarily novel functional networks in the human brain? *J Neurosci*. 2013; 33:3259–3275.10.1523/JNEUROSCI.4392-12.2013 [PubMed: 23426655]
- Mantini D, Hasson U, Betti V, Perrucci MG, Romani GL, Corbetta M, Orban GA, Vanduffel W. Interspecies activity correlations reveal functional correspondence between monkey and human brain areas. *Nature Publishing Group*. 2012; 9:277–282.10.1038/nmeth.1868
- Maunsell JH, Nealey TA, DePriest DD. Magnocellular and parvocellular contributions to responses in the middle temporal visual area (MT) of the macaque monkey. *J Neurosci*. 1990; 10:3323–3334. [PubMed: 2213142]
- Maunsell JH, Van Essen DC. Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. *J Neurophysiol*. 1983; 49:1127–1147. [PubMed: 6864242]
- Nassi JJ, Callaway EM. Parallel processing strategies of the primate visual system. *Nat Rev Neurosci*. 2009; 10:360–372.10.1038/nrn2619 [PubMed: 19352403]
- Nelissen K, Vanduffel W, Orban GA. Charting the lower superior temporal region, a new motion-sensitive region in monkey superior temporal sulcus. *J Neurosci*. 2006; 26:5929–5947.10.1523/JNEUROSCI.0824-06.2006 [PubMed: 16738235]
- Oram MW, Perrett DI. Integration of form and motion in the anterior superior temporal polysensory area (STPa) of the macaque monkey. *J Neurophysiol*. 1996; 76:109–129. [PubMed: 8836213]
- Orban, GA.; Jastorff, J. Functional Mapping of Motion Regions in Human and Nonhuman Primates. In: Werner, JS.; Chalupa, LM., editors. *The New Visual Neurosciences*. MIT Press (MA); Cambridge: 2013. p. 777-791.
- Parr LA. The evolution of face processing in primates. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2011; 366:1764–1777.10.1098/rstb.2010.0358
- Perrett DI, Smith PAJ, Potter DD, Mistlin AJ, Head AS, Milner AD, Jeeves MA. Visual Cells in the Temporal Cortex Sensitive to Face View and Gaze Direction. *Proceedings of the Royal Society of London Series B, Biological Sciences*. 1985; 223:293–317.
- Polosecki P, Moeller S, Schweers N, Romanski LM, Tsao DY, Freiwald WA. Faces in motion: selectivity of macaque and human face processing areas for dynamic stimuli. *J Neurosci*. 2013; 33:11768–11773.10.1523/JNEUROSCI.5402-11.2013 [PubMed: 23864665]
- Popivanov ID, Jastorff J, Vanduffel W, Vogels R. Heterogeneous single-unit selectivity in an fMRI-defined body-selective patch. *J Neurosci*. 2014; 34:95–111.10.1523/JNEUROSCI.2748-13.2014 [PubMed: 24381271]
- Puce A, Perrett D. Electrophysiology and brain imaging of biological motion. *Philos Trans R Soc Lond, B, Biol Sci*. 2003; 358:435–445.10.1098/rstb.2002.1221 [PubMed: 12689371]
- Saleem, KS.; Logothetis, NK. *A Combined MRI and Histology Atlas of the Rhesus Monkey Brain in Stereotaxic Coordinates*. Academic Press; 2012.
- Schein SJ, de Monasterio FM. Mapping of retinal and geniculate neurons onto striate cortex of macaque. *J Neurosci*. 1987; 7:996–1009. [PubMed: 3033167]

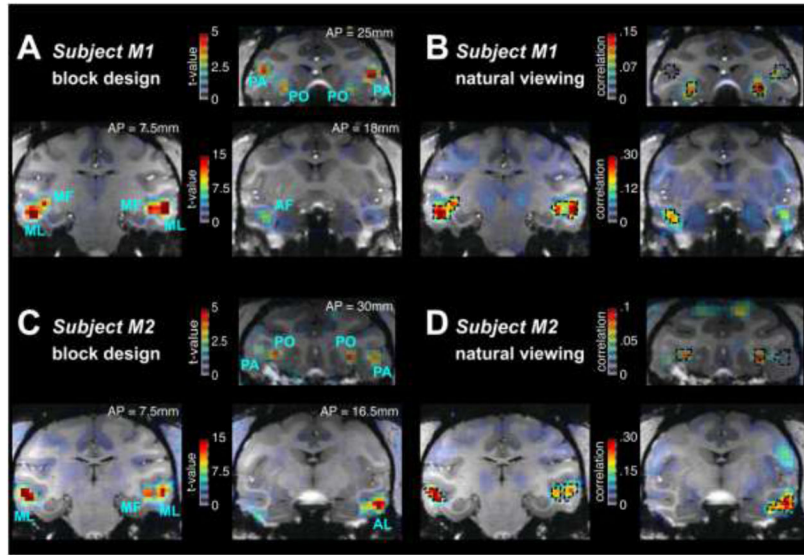
- Schiller PH, Logothetis NK, Charles ER. Role of the color-opponent and broad-band channels in vision. *Vis Neurosci.* 1990; 5:321–346. [PubMed: 2265148]
- Scholte HS, Ghebreab S, Waldorp L, Smeulders AWM, Lamme VAF. Brain responses strongly correlate with Weibull image statistics when processing natural images. *JOV.* 2009; 9(29):1–15.10.1167/9.4.29
- Sheinberg DL, Logothetis NK. Noticing familiar objects in real world scenes: the role of temporal cortical neurons in natural vision. *J Neurosci.* 2001; 21:1340–1350. [PubMed: 11160405]
- Shepherd SV, Steckenfinger SA, Hasson U, Ghazanfar AA. Human-monkey gaze correlations reveal convergent and divergent patterns of movie viewing. *Curr Biol.* 2010; 20:649–656.10.1016/j.cub.2010.02.032 [PubMed: 20303267]
- Singer JM, Sheinberg DL. Temporal cortex neurons encode articulated actions as slow sequences of integrated poses. *J Neurosci.* 2010; 30:3133–3145.10.1523/JNEUROSCI.3211-09.2010 [PubMed: 20181610]
- Tsao DY, Freiwald WA, Knutsen TA, Mandeville JB, Tootell RBH. Faces and objects in macaque cerebral cortex. *Nat Neurosci.* 2003; 6:989–995.10.1038/nn1111 [PubMed: 12925854]
- Tsao DY, Freiwald WA, Tootell RBH, Livingstone MS. A cortical region consisting entirely of face-selective cells. *Science.* 2006; 311:670–674.10.1126/science.1119983 [PubMed: 16456083]
- Tsao DY, Moeller S, Freiwald WA. Comparing face patch systems in macaques and humans. *Proceedings of the National Academy of Sciences.* 2008a; 105:19514–19519.10.1073/pnas.0809662105
- Tsao DY, Schweers N, Moeller S, Freiwald WA. Patches of face-selective cortex in the macaque frontal lobe. *Nat Neurosci.* 2008b; 11:877–879.10.1038/nn.2158 [PubMed: 18622399]
- Ungerleider LG, Bell AH. Uncovering the visual “alphabet”: advances in our understanding of object perception. *Vision Res.* 2011; 51:782–799.10.1016/j.visres.2010.10.002 [PubMed: 20971130]
- Ungerleider LG, Desimone R. Cortical connections of visual area MT in the macaque. *J Comp Neurol.* 1986; 248:190–222.10.1002/cne.902480204 [PubMed: 3722458]
- Xiang QS, Ye FQ. Correction for geometric distortion and N/2 ghosting in EPI by phase labeling for additional coordinate encoding (PLACE). *Magn Reson Med.* 2007; 57:731–741.10.1002/mrm.21187 [PubMed: 17390358]

**HIGHLIGHTS**

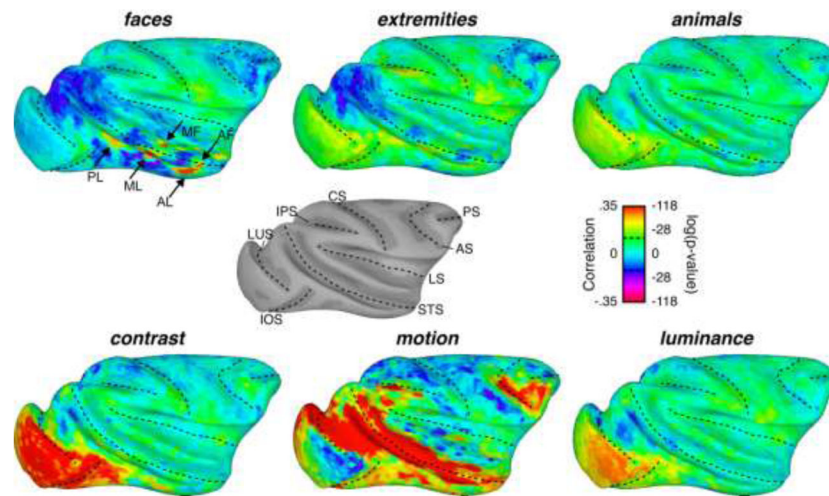
- We used a natural viewing paradigm to generate fMRI feature maps.
- We were able to map face patches with as little as 15 minutes of natural viewing.
- Motion dominated fMRI responses throughout the ventral visual pathway.
- Motion contributed much less to fMRI responses to videos without social content.



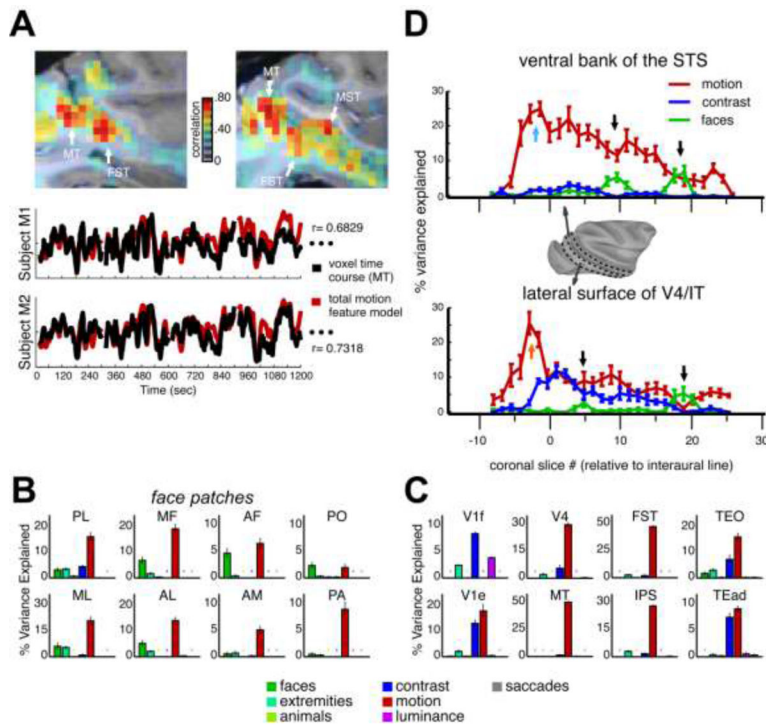
**Figure 1.**  
 A) Schematic of the single session of the monkeys' participating in the natural viewing paradigm. B) Time courses of a voxel, from area MT of subject M2, and the individual features used in the correlation and step-wise regression analyses. Time courses have been concatenated across the 15 social movies presented throughout the experiment. Feature time courses are in arbitrary units following down-sampling and being convolved with a MION function (see **METHODS**). C) The correlation matrix observed between each feature time courses shown in B.



**Figure 2.**  
 A) Activation maps from using conventional block design to map face responsive regions of the cortex of subject M1. Blue labels show the location of face patches as defined using the Tsao and colleagues designations (Tsao et al., 2008b; 2003). B) Activation maps exhibit the correlation between each voxel and the face feature model from M1. The overlaid dotted black lines show the location of the face patches found using the conventional block design in A. **Figures C & D** show the location of the face patches based on conventional block design (C) and natural viewing paradigm (D) for subject M2. Color maps use a graded transparency to depict the functional data overlaying the anatomy. Increasing t-values (A & C) or correlations (B & D) are shown with greater opacity; no threshold was applied to the data. *Face Patch designation.* MF: middle fundus; ML: middle lateral; AF: anterior fundus; PA: prefrontal arcuate (area 45); PO: prefrontal orbital (area 131); PL: posterior lateral. Anterior-Posterior (AP) coordinates of the individual coronal slices relative to the inter-aural slice of each subject are displayed in A and C.



**Figure 3.** Inflated surface maps exhibiting the correlation between a given feature time course and the average brain activity in response to 15 movies. The dotted black lines on the surface maps represent the locations of the fundi of major sulci on the macaque brain. The gray-scale surface map depicts the surface anatomy from subject M2, who's data is depicted here, with labels for the various sulci (LUS: lunate sulcus; iOS: inferior occipital sulcus; IPS: intraparietal sulcus; CS: central sulcus; STS: superior temporal sulcus; LS: lateral sulcus; AS: arcuate sulcus; PS: principal sulcus). Black arrows in the *faces* surface map show the approximate locations of the face patches (see Figure 2 legend for label names). Dotted lines in the scale bar represent a significant correlation at  $p=0.05$ , Bonferroni corrected.

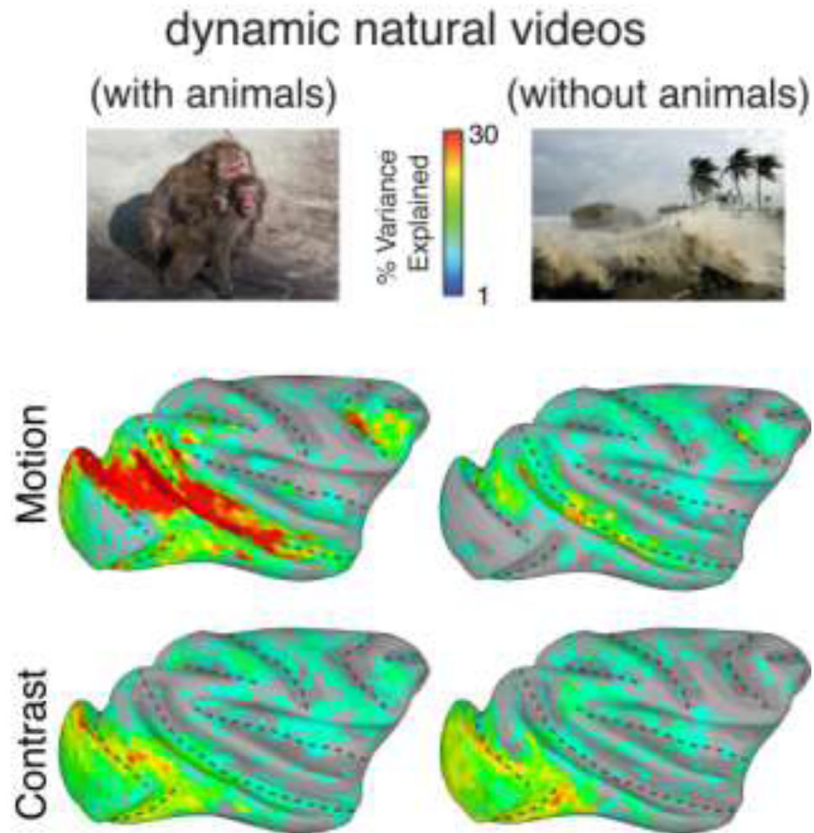


**Figure 4.**

A) *Top.* Sagittal views of motion sensitive regions in subject M2. Data represent the correlation between each voxel’s response to 15 movies and the motion feature model. White arrows show the approximate location of areas MT, FST and MST within these slices of the right hemisphere. *Bottom.* Time courses show the similarity between the *motion* feature model from 4 movies (red) and the time course in response to those movies from one voxel in area MT (black) from M1 (top) and M2 (bottom). Figures B & C) Bar graphs showing the relative percent variance explained by each of the seven feature models submitted to the stepwise regression for 16 ROIs in the right hemisphere of M2. Colored stars show that for that ROI the given feature model was not included in any of the voxels used because it did not add an additional 1% explained variance on top of the other models. Error bars represent the standard error of the mean variance explained for each feature model across the voxels used in the ROI. B) The relative percent variance explained in eight face patches. C) The percent variance explained within visually responsive regions outside of the face patches. Results were similar across the 3 monkeys. *ROI labels.* PL: posterior lateral; ML: middle lateral; MF: middle fundus; AL: anterior lateral; AF: anterior fundus; AM: anterior medial; PA: PO: prefrontal orbital (area 131); prefrontal arcuate (area 45); V1f: foveal primary visual cortex; V1e: extrafoveal V1; V4: visual area 4; MT: middle temporal area; FST: floor of superior temporal area; IPS: interparietal sulcus; TEO: posterior lateral inferior temporal cortex; TEad: dorsal subregion of anterior temporal lobe. D) The average percent variance explained by the motion (red), contrast (blue), and face (green) feature models from posterior to anterior in the ventral bank of the STS (top) and lateral surface of V4/IT (bottom). X-axis shows the coronal slice number relative to the interaural plane. Black arrows show the relative position of 4 of the face patches in Figures 2C & D. Blue and orange arrows show the relative position of areas MT and V4 in the ventral bank and lateral

surface plots, respectively. Error bars represent the standard error of the mean of the percent variance explained by each feature within each coronal slice. The surface map from the right hemisphere of M2 shows the boundaries of the two regions of interest used. Posterior to anterior gradient was found in all three monkeys.





**Figure 5.**

*Left Column* Surface maps showing the percent variance explained by the motion (top) and contrast (bottom) feature models to natural movies that contained animals interacting with each other and the environment. *Right Column.* Surface maps showing the percent variance explained by the motion (top) and contrast (bottom) feature models to natural movies that contained *no* animals. Reduced explained variance for motion in the nonsocial movies was found in both subjects tested.

**Table 1**

Face patches identified by the block and natural viewing paradigms within each hemisphere for all 3 animals. Italicized and underlined subjects indicate instances in which only one of the testing regimes identified the particular face patch.

Face Patch	Left Hemisphere		Right Hemisphere	
	Block Design	Natural Viewing	Block Design	Natural Viewing
Posterior Lateral (PL)	M1, M2	M1, M2, <u>M3</u>	M1, M2	M1, M2
Middle Lateral (ML)	M1, M2, M3	M1, M2, M3	M1, M2, M3	M1, M2, M3
Middle Fundus (MF)	M1, M2, M3	M1, M2, M3	M1, M2, M3	M1, M2, M3
Anterior Lateral (AL)	M2, M3	M3	M1, M2, M3	M1, M2, M3
Anterior Fundus (AF)	M1, M2	M1, M2	M1, M2, M3	M1, M2, M3
Anterior Medial (AM)	M1, M2	M1, M2, <u>M3</u>	M2	<u>M1</u> , M2, <u>M3</u>
Prefrontal Orbital (PO)	M1, M2	M1, M2	M1, M2	M1, M2
Prefrontal Arcuate (PA)	<u>M1</u>		M1, M2	M1, M2