NeuroImage 269 (2023) 119907

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

The contribution of dynamics to macaque body and face patch responses

A. Bognár[a,b,1], R. Raman[a,b,1], N. Taubert[c], Y. Zafirova[a,b], B. Li[d], M. Giese[c], B. De Gelder[d,e], R. Vogels[a,b,*]

[a] Deparment of Neurosciences, KU Leuven, Leuven, Belgium
[b] Leuven Brain Institute, KU Leuven, Leuven, Belgium
[c] Department of Cognitive Neurology, University of Tuebingen, Tuebingen, Germany
[d] Department of Cognitive Neuroscience, Maastricht University, Maastricht, the Netherlands
[e] Department of Computer Science, University College London, London, UK

ARTICLE INFO

Keywords:
Faces
Bodies
Inferior temporal cortex
Action
Monkey fMRI

ABSTRACT

Previous functional imaging studies demonstrated body-selective patches in the primate visual temporal cortex, comparing activations to static bodies and static images of other categories. However, the use of static instead of dynamic displays of moving bodies may have underestimated the extent of the body patch network. Indeed, body dynamics provide information about action and emotion and may be processed in patches not activated by static images. Thus, to map with fMRI the full extent of the macaque body patch system in the visual temporal cortex, we employed dynamic displays of natural-acting monkey bodies, dynamic monkey faces, objects, and scrambled versions of these videos, all presented during fixation. We found nine body patches in the visual temporal cortex, starting posteriorly in the superior temporal sulcus (STS) and ending anteriorly in the temporal pole. Unlike for static images, body patches were present consistently in both the lower and upper banks of the STS. Overall, body patches showed a higher activation by dynamic displays than by matched static images, which, for identical stimulus displays, was less the case for the neighboring face patches. These data provide the groundwork for future single-unit recording studies to reveal the spatiotemporal features the neurons of these body patches encode. These fMRI findings suggest that dynamics have a stronger contribution to population responses in body than face patches.

1. Introduction

fMRI studies in humans and monkeys demonstrated body-category selective areas in the occipito-temporal lobe (Peelen and Downing, 2007) that are activated stronger by images of bodies compared to objects and faces. In the macaque, several studies have mapped body patches using static images (Bao et al., 2020; Bell et al., 2009, 2011; Fisher and Freiwald, 2015b; Lafer-Sousa and Conway, 2013; Pinsk et al., 2005; Popivanov et al., 2012; Premereur et al., 2016; Sliwa and Freiwald, 2017; Tsao et al., 2003). Across the different stimulus sets and contrasts used in these studies, the most consistently observed body patches are two in the lower bank of the Superior Temporal Sulcus (STS; for review see Vogels (2022)). The middle STS body (MSB) patch (Kumar et al., 2017; Popivanov et al., 2014) is located anterior and lateral to the Fundus of the Superior Temporal (FST) area and was observed in several monkey fMRI mapping studies (Bao et al., 2020; Bell et al., 2009, 2011; Fisher and Freiwald, 2015b; Lafer-Sousa and Conway, 2013; Pinsk et al., 2005; Popivanov et al., 2012; Premereur et al., 2016; Sliwa and Freiwald, 2017; Tsao et al., 2003). The second one, the anterior STS body (ASB) patch (Kumar et al., 2017), is located anteriorly in the lateral part of the lower bank of the STS and was also evident in most studies (Bao et al., 2020; Bell et al., 2009, 2011; Fisher and Freiwald, 2015b; Lafer-Sousa and Conway, 2013; Pinsk et al., 2009, 2005; Popivanov et al., 2012; Premereur et al., 2016; Sliwa and Freiwald, 2017). Recent studies (Bao et al., 2020; Premereur et al., 2016; Sliwa and Freiwald, 2017) have also observed a smaller body patch more anteriorly and ventral in the inferior temporal (IT) cortex, which we labeled as the Anterior Ventral Body (AVB) patch (Vogels, 2022). Single-unit recordings in these body patches showed a high proportion of neurons that responded stronger to images of bodies (of e.g. monkeys, humans, four-legged mammals and birds) than to faces, vegetables, fruits and manmade objects (Bao et al., 2020; Bao and Tsao, 2018; Bell et al., 2011; Kumar et al., 2017; Popivanov et al., 2014; Vogels, 2022).

These monkey fMRI mapping studies of body patches employed static images of bodies. However, given the important role of dynamics for action recognition (Giese and Poggio, 2003), including signaling

* Corresponding author at: Deparment of Neurosciences, KU Leuven, Leuven, Belgium.
E-mail address: rufin.vogels@kuleuven.be (R. Vogels).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.neuroimage.2023.119907.
Received 7 September 2022; Received in revised form 20 December 2022; Accepted 26 January 2023
Available online 27 January 2023.

emotion (de Gelder et al., 2015; Grezes et al., 2007), the body patch network might be more extensive when mapped with dynamic displays of bodies. Also, the monkey fMRI studies using static stimuli observed activations to bodies, relative to objects, in the upper bank of the STS, but this was inconsistent among subjects (Popivanov et al., 2012) and activations to bodies were similar to those to faces in the upper bank (Jastorff et al., 2012). Because upper bank STS neurons have been reported to show a greater response to moving compared to static stimuli (Baylis et al., 1987; Pitcher and Ungerleider, 2021; Vangeneugden et al., 2011, 2009), dynamic body displays may robustly reveal body-related activations in the upper bank of the STS, thus extending the body processing network.

In the present study, we mapped with fMRI body patches in macaque monkeys using a novel stimulus set that consisted of displays of naturally acting macaque monkeys. We contrasted the response to the monkey body movements with monkey faces showing expressive displays and/or head rotations and with moving artificial objects. To isolate activations to bodies, we blurred the faces in the monkey body displays. In addition, we measured activations to two types of scrambled displays to control for low-level spatiotemporal features. In an additional experiment, we compared the activations to the dynamic displays with those to static images sampled from the same dynamic displays, assessing the contribution of dynamics to the activation.

Using the dynamic stimulus sets, we observed a larger number of body patches than those reported before with static stimuli, in particular in the upper bank of the STS. In addition, we found, with identical stimulus displays in the same scans, a stronger contribution of dynamics in the body patches than in the neighboring face patches, suggesting a stronger role of motion in driving the responses in body than in face patches.

2. Methods

2.1. Subjects

Five rhesus monkeys (*Macaca mulatta*) contributed to this study. The monkeys were housed in pairs or triplets. The monkeys were implanted with a plastic headpost, using ceramic screws and dental cement following standard aseptic procedures and full anesthesia. They were trained to fixate continuously a small target point for juice rewards in a mock scanner set up for several weeks before scanning. Animal care and experimental procedures complied with the regional (Flanders) and European guidelines and were approved by the local Animal Ethical Committee.

2.2. Stimuli

The 20 body stimuli were taken from footage of rhesus monkeys from our colony which were filmed when they were expressing natural behavior in interaction with another monkey or a human in their large enclosure. The walls of the enclosure, except for the windows, were covered by a green sheet. This allowed the segmentation of the monkey from the background of the videos. We selected 20 snippets of 1 s duration in which a single-acting monkey was visible. The selected actions included grasping, picking, turning, walking, threatening, throwing, wiping, and initiating jumping. The displayed actions depicted a large variety of poses and body orientations, even within a single movie, which should facilitate the activation of body patches. The face of the monkey was blurred so that facial expression and identity were unrecognizable. The translational component of the movement of the monkey across the display, when present (e.g. during walking) was removed and the monkey's body was centered. Each movie was resized so that the maximal extent of the re-centered monkey's body (including its head) fitted in a 6 by 6 deg square.

The 20 face videos were taken from two footages of rhesus monkeys' faces. The first footage was obtained from a published comparative study of facial expressions (Zhu et al., 2013). Twelve snippets of 1 s that showed frontal face movements such as chewing, lip-smacking, fear grin, and threat were selected. The face was cut out and centered. The other 8 movies of faces were based on the same footage as that of the bodies. Snippets of 1 s duration in which the head of the monkey was moving, e.g. rotating from frontal to profile view, were selected and the head was segmented from the rest of the body and background. The head was centered. As for the bodies, the face movies were resized so that the maximal extent of the dynamic face fitted in a 6 by 6 deg square.

The 20 object stimuli included computer-rendered objects that depicted movement. Movies of a 1 s duration of a variety of moving objects were employed: multi-part mechanical objects (taken from https://garethwashere.tumblr.com), of which the parts made different non-rigid movements, a rotating airplane, and cars with different motion patterns (e.g. rocking or jumping). The moving objects were centered and resized so that their maximal extent fitted in a 6 by 6 deg square.

Because the color of monkeys' fur has a limited range and differs from the face and many objects, we rendered achromatic versions of the original color movies. The use of achromatic movies facilitated the comparison of the activation to objects and bodies, eliminating systematic differences in color composition between the two categories. The movies were rendered with a 60 Hz frame rate. The bodies, faces, and objects were presented on top of a dynamic white noise background (Fig. 1). The background consisted of single pixels of which the gray level was randomly sampled from a uniform distribution at a rate of 30 Hz.

We examined whether the videos of the three categories differed in low-level spatial and temporal features. The broad distributions of the luminance (**Fig. S1**), pooled across all frames of all movies per category, overlapped strongly among the three categories. The spatial frequency spectra (**Fig. S1**), averaged across all frames and movies, showed some minor differences among the categories, with the difference in mean power among the categories depending on spatial frequency (e.g. the expected higher power at low but not high spatial frequencies for the faces). We estimated the local motion energy of each movie using the Lucas Kanade derivative of Gaussian filter optical flow algorithm (as in Zhang et al. (2020)) implemented by the opticalFlowLKDOG Matlab function. The distributions of the estimated motion energy computed using the 6 by 6 deg canvas including the dynamic noise, overlapped among the three categories. The face stimuli had a lower estimated local motion energy than the objects and bodies, while there was no significant difference between objects and bodies (**Fig. S2**).

To control for low-level feature differences among the three categories, we included two additional sets of stimulus conditions: mosaic-scrambled (experiment 1) and phase-scrambled movies (experiment 2). For each of the 60 original movies, we constructed mosaic-scrambled versions (Fig. 1). The 6 by 6 deg square (210×210 pixels) that consisted of the dynamic body, face, or object, was divided into 0.4 by 0.4 deg blocks and these were randomly replaced within 5 neighboring horizontal bands of 1.2 deg in height. The latter ensured that the local motion, luminance, and contrast distribution of the original movie was kept in the scrambled movie for the lower, middle, and upper visual field. Especially for some of the face stimuli, movement differed between the lower (mouth region) and upper visual field, and we wanted to keep this difference in local motion between the lower and upper visual field also in the mosaic scrambled control stimuli. The mosaic scrambling impaired the whole shape and global motion of the dynamic bodies, faces, and objects, but maintained local motion. The mosaic-scrambled stimuli had the same luminance, contrast, and non-background area as the original movies. The mosaic-scrambled movies were presented to 3 monkey subjects (experiment 1).

Some of the body fragments, e.g. part of the hands/fingers in the case of bodies and part of an eye in the case of the faces, were still recognizable on close inspection in the mosaic-scrambled stimuli. To assess how naïve observers would categorize the mosaic-scrambled movies we

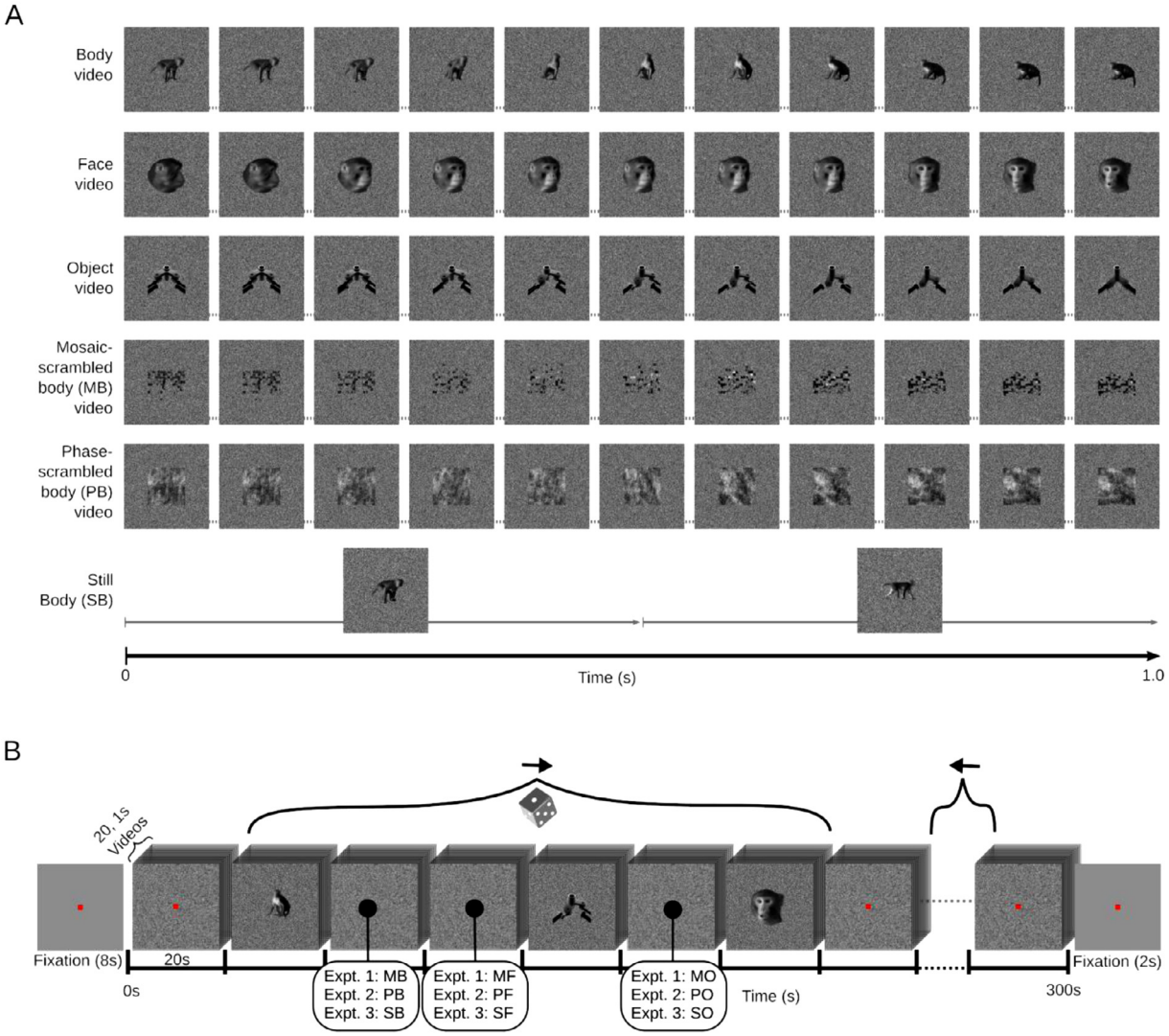

Fig. 1. Stimulus displays and design. **A.** Frames of an example video of each stimulus condition. Each row shows frames that are 100 ms apart in the 1 s movie in the presented order. The still images were shown for 500 ms each back to back. **B.** Illustration of an example run of the block design. The run started and ended with a display with only a fixation target that the animal was required to fixate. Blocks lasted 20 secs and consisted of 20 videos of 1 s each or 40 static images of 500 ms each. Stimuli were presented back-to-back. The order of the stimulus conditions was randomized across runs. The 6 stimulus conditions were presented twice in opposite order in-between 3 fixation conditions in which only the noise background and fixation target were presented. Although not shown in the figure, the fixation target was present in all conditions throughout the run. The 3 experiments differed with respect to the scrambled or static conditions (Experiment 1: dynamic mosaic-scrambled (M*); Experiment 2: dynamic phase-scrambled (P*); Experiment 3: static (S*) stimuli). MB: mosaic-scrambled bodies; PB: phase-scrambled bodies; SB: static bodies; MF: mosaic-scrambled faces; PF: phase-scrambled faces; SF: static faces; MO: mosaic-scrambled objects; PO: phase-scrambled objects; SO: static objects.

showed these movies to 12 human subjects (5 female; mean age: 31 years; age range 21 - 51 years) who did not see the original, unscrambled movies and were unaware of their content (5 subjects saw only 50% of the movies for technical reasons). The 60 mosaic-scrambled movies together with 60 mosaic-scrambled movies of human bodies, human faces, and another set of objects were presented and the subjects were instructed to categorize each of these using 3 classes: face/head, animal/human, and object (level 1 classification). Then, based on their category choice, they were presented with 5 subcategories (for face/head: human face, monkey face, horse head, bird head, cat head; for animal/human: human, monkey, cat, horse, and bird) of which they had to choose one (level 2 classification). For the mosaic-scrambled monkey body and face movies, the percent of animate category choices was greater than the object choices and there was a tendency for a higher percentage of the face and body choices for the face and body movies, respectively (level 1 classification; **Fig. S3**). The mosaic-scrambled object movies were categorized mainly as objects and animals. Thus, a rough superordinate animate/inanimate categorization of the mosaic-scrambled movies was possible. The level-2 classification (ordinate level) showed however that the human subjects were unable to identify the mosaic-scrambled monkey body and face movies as depicting monkeys.

The second control employed Fourier phase scrambling (Fig. 1), which abolished shape information but maintained the global motion. We performed a 2D Fourier transform of each frame of the original movie and then applied the same random rotation of the phase spectrum for all the frames of the movie, before performing the inverse Fourier transform (as in Sliwa and Freiwald (2017)). This procedure kept the spatial frequency spectrum, mean luminance, and global motion of the original movies. The phase scrambled movies were presented to two monkeys (experiment 2; other animals than the 3 shown the mosaic-scrambled movies).

To the same two monkeys, we presented also conditions with static frames of each of the 60 movies (experiment 3). We selected 2 frames for each movie that depicted different poses or object views. The 40 static stimuli of a category were presented for a duration of 500 ms each in random order (Fig. 1). The white noise background was static in the static stimulus conditions.

2.3. fMRI scanning procedure and design

During scanning, the monkeys sat in a horizontal sphinx position with their heads fixed in an MRI-compatible chair. The chair was positioned in front of a translucent screen, at a distance of approximately 58 cm. The stimuli were projected (Barco 6300 LCD projector) on the screen. Eye position was continuously monitored (120 Hz; Iscan) and the animals were performing a fixation task (fixation window size approximately 2–3 deg; (Popivanov et al., 2012)) during scanning for a juice reward.

The monkeys were scanned with a 3T Siemens Trio scanner following standard procedures (Vanduffel et al., 2001). Functional MRI images were acquired using a custom-made 8-channel monkey coil, a saddle-shaped radial transmit-only surface coil (Ekstrom et al., 2008; Kolster et al., 2009), and a gradient-echo T2∗-weighted echo-planar imaging sequence of 40 slices and flip angle of 90° (repetition time TR = 2000 ms, echo time TE = 18 ms, 1.25 mm isotropic voxel resolution). Slices were oriented transversally covering the whole brain. We obtained high-resolution anatomical MRI images for each monkey in a separate session under Ketamine/ Medetomidine anesthesia, using a single radial transmit–receive surface coil and a Magnetization-Prepared Rapid Acquisition with Gradient Echo (MPRAGE) sequence (TR = 2700 ms, TE = 3.5 ms, flip angle = 9°, 208 slices, 0.4 mm isotropic voxel resolution). To increase the signal-to-noise ratio (Leite et al., 2002), we injected the contrast agent Monocrystalline Iron Oxide Nanoparticle (MION; Molday ION; 8–11 mg/kg) into the femoral/saphenous vein immediately before scanning.

We employed block designs in all experiments. The stimuli of a category were presented in a block back to back in random order. Thus, one block, consisting of 20 movies or 40 static stimuli, lasted 20 s (Fig. 1). The order of the stimuli within a block was randomized across runs. There was a period of fixation of the fixation target on an otherwise empty screen at the beginning (8 s; not included in the analysis) and the end (2 s) of a run. This period was followed by a fixation block in which the fixation target (size = 0.2 deg) was shown together with the dynamic white noise background for 20 secs. Then, 6 blocks of 6 different conditions were presented (experiments 1 & 2: dynamic bodies, dynamic faces, dynamic objects, scrambled bodies, scrambled faces, scrambled objects; experiment 3: dynamic bodies, dynamic faces, dynamic objects, static bodies, static faces, and static objects) followed by a noise background fixation block. After this fixation block, we repeated the same 6 blocks but in the opposite order (palindromic sequence), followed by a noise background fixation block. The fixation target was present for all stimulus conditions. The order of the 6 blocks was randomized across runs using a balanced Latin square design (Fig. 1).

Three monkeys (O, N & G) were scanned using the original and mosaic scrambled movies (experiment 1). The other 2 monkeys (J & H) participated in experiments 2 and 3. In experiment 2, we presented the original and the phase scrambled movies, while in experiment 3, we showed the original movie and static stimulus conditions.

2.4. Data analysis: imaging

Percent fixation duration was computed online during scanning using an algorithm which included eye position deviations of less than 200 ms (blinks) outside the fixation window. Only runs in which the monkeys were fixating at least 89% during the presentation of the 6 stimulus conditions and the noise background fixation condition were included in the analysis. For pre-processing, we re-oriented the images and applied slice timing correction in SPM. A non-rigid, slice-by-slice realignment within runs and affine realignment between runs within a day was performed for motion-correction with jip-align (https://www.nmr.mgh.harvard.edu/~jbm/jip). Mean functional images were then non-rigidly co-registered to the T1 anatomical images of the same monkey in SPM12. Images were smoothed in SPM12 with an isotropic Gaussian kernel (FWHM: 1.5 mm).

Subsequent data analysis was performed with SPM12. All valid runs were combined in a fixed-effects model for each subject separately in their native space. They were analyzed with a general linear model (GLM) with 7 regressors (6 stimulus conditions + fixation condition), plus 9 additional head-motion (6 covariates: translation and rotation in 3 dimensions) and eye-movement regressors (2 covariates for the horizontal and vertical eye position, and one for the pupil diameter) per run. Each condition was modeled using a convolution with a Gamma function (delta = 0, tau = 8 and exponent = 0.3), modeling the MION hemodynamic response function.

We defined body patches with different, complementary sets of contrasts. In each case, the resulting t-maps were thresholded at $p = 0.05$, Family-Wise Error (FWE) rate, corresponding to $t = 4.9$. In the first set of contrast, we defined body patches using an inclusive mask of bodies – faces, and bodies – objects, revealing those voxels that were activated more by bodies compared to faces and objects. Note that this is more stringent than the contrast bodies – (faces + objects) since in the latter case bodies and for instance, faces may still activate equally since the activations to faces are averaged with those to the objects. In the second set of contrasts, we included a third contrast, bodies – scrambled bodies, using inclusive masking with the other two contrasts. This controlled for the contribution of low-level stimulus features to the body activations. Other contrasts are described in the relevant Results sections and all contrasts employed for the main analyses are listed in Table S1.

For display purposes, we show the activations for the different experiments on flat maps. To do this, we first co-registered the individual monkey data to a template anatomy (M12, Ekstrom et al. (2008)) and then used Caret software to present the activations on the flattened 3D image of the template. Note that flat maps provide an illustration of the relative location of the patches but suffer from distortions and small patches may not be visible. That is why we stress that the coronal sections in native space that we provide in the main figures and not the flat maps, should be taken to localize patches.

In monkeys J and H, we defined body and face patch Regions Of Interests (ROIs) using the data of experiment 2. To do so, we defined body patch ROIs using the inclusive masking of bodies – objects, bodies – faces, and bodies – scrambled bodies, while face patch ROIs were defined with the inclusive masking of faces – objects and faces – scrambled faces. We did not include the faces – bodies contrast, because it was possible that our body stimuli, which have blurred faces, also activate face patches (Cox et al., 2004; Fisher and Freiwald, 2015b). Using the SPM MarsBaR toolbox, the ROI was built using the box option, with the center being the local maximum of the patch and the width being 2. Thus, the ROI consisted of 8 voxels including the local maximum. For each of the ROIs defined using the data of experiment 2, we then computed with MarsBaR the mean percent signal change (PSC) for each of the conditions, relative to the baseline white-noise fixation condition, for each of the runs of experiment 3. For further statistical analyses, we took the average of the PSC of the homotopic patches of the two hemispheres. For each ROI (pooled across hemispheres) and monkey, we performed a repeated-measures ANOVA on the PSCs with factors stimulus category (bodies; faces; object) and static versus dynamic. For each body patch ROI, we normalized the PSC for each category by dividing it by the PSC for the dynamic bodies. Likewise, for each face patch ROI, we normalized the PSCs by dividing these by the PSC for the dynamic faces. To compare the effect of motion for the same stimuli between the face and body patches, we subtracted the normalized PSC for the static stimuli from the normalized PSC for the dynamic stimuli. This was done after pooling the normalized PSC data of the two monkeys, after equating

the number of runs across the monkeys. 95% confidence intervals of the mean normalized PSC and of the mean difference in normalized PSC were computed by bootstrapping runs. To do so, we resampled with replacement the percent signal change of the runs and for each set of resampled runs, we computed the mean percent signal change (mean across the resampled runs). The resampling was done 1000 times, generating 1000 means. The 95% confidence interval was defined as the 2.5 and 97.5 percentile of the distribution of the 1000 means. In addition, we computed for each run the difference in percent signal change between the two relevant conditions and then resampled these differences, i.e. the runs, 1000 times followed by computation of the 95% confidence interval using the percentile method.

2.5. Data analysis: eye movements

To assess whether putative differences in eye movements among the stimulus conditions can explain the fMRI activations, we analyzed the eye movement data obtained during the fMRI scans of our 5 monkeys in the 3 experiments. As a first step, we averaged the mean percent fixation (within the fixation window) for each condition separately for each of the monkeys and experiments for the same runs that entered the fMRI analysis. Second, we performed the following analysis of the eye movements of the same runs. The eye movement data, saved with a 1 kHz resolution, were down-sampled to 120 Hz, which is the sampling rate of the eye tracker employed in the scanner. Then we aligned the data across runs by subtracting per run the mean eye movement position of the noise background fixation condition (in which no images or videos were presented) from the eye movement traces of all conditions. Subsequently, we removed blinks and large saccades that occurred outside a window of 16 deg (centered on the fixation target; note that the stimuli were 6 deg in size) using a position threshold on the horizontal and vertical traces. In addition, the eye signal 50 ms before and after a blink or large eye movement was also removed. The remaining "valid" eye signal data were concatenated for each condition and run and employed for further eye movement analysis. The percent of "valid" eye movement data was computed for each condition and is shown in **Fig. S4**. Note that although there are numerical differences among the stimulus conditions in the percent of kept data for the subsequent eye movement analysis, these are small and inconsistent across monkeys. We employed the method of Vergnieux and Vogels (2020), based on the algorithm of Engbert and Kliegl (2003), to detect saccades. After filtering the eye movement trace using a 40 Hz low-pass filter (5th order Butterworth), horizontal and vertical eye velocities (Engbert and Kliegl, 2003; Vergnieux and Vogels, 2020) were computed. Saccades were detected using an elliptic threshold with a linear factor lambda (Engbert and Mergenthaler, 2006) of 4 in velocity space. Moreover, only eye movements faster than 10 deg/s and larger than 0.3 deg were taken as saccades. The duration of the saccade had to be at least 16.7 ms and the interval between two saccades at least 50 ms (Kaliukhovich and Vogels, 2011), otherwise, the second putative saccade was discarded in favor of the first. These criteria were implemented to reduce false positive saccade detections. We computed the saccade frequency, direction, and amplitude for each condition separately. In addition to saccades we also analyzed the instantaneous eye velocities since these may have been affected by the dynamic stimuli. To reduce noise following differentiation of the signal, we smoothed the computed velocity traces using a Gaussian filter with a standard deviation of 16.7 ms. The distribution of the instantaneous eye speed for each monkey and experiment were computed per condition for the horizontal and vertical direction separately.

3. Results

3.1. Dynamic body patches in the visual temporal cortex

We mapped dynamic body patches by inclusively masking two contrasts: dynamic bodies minus dynamic faces and dynamic bodies minus dynamic objects in experiments 1 and 2. The body patch activations are shown for each monkey separately on coronal sections in native space (Figs. 2 and 3) and as flat maps in the Supplemental Materials (**Fig. S5**). Note that some activations that were present in the coronal sections were not visible on the flat maps and flat maps suffer from spatial distortions. Thus, we will rely on the coronal sections to describe the patches. The most posterior body-selective activations were present in posterior STS motion areas: MT (in the ventral STS) and FST (in the fundus of the STS). The FST activation was present in each of the 5 monkeys (experiment 1: Fig. 2 and experiment 2: Fig. 3) and we will label this patch as the posterior STS body patch (PSB). MT and FST are typically observed in monkey fMRI studies of actions (Nelissen et al., 2006). Located more anterior and medial in the lower bank of the STS, the middle STS body patch (MSB) was activated by the dynamic body images. MSB is the most consistently observed body patch when using static stimuli (Bao et al., 2020; Bao and Tsao, 2018; Kumar et al., 2017; Popivanov et al., 2012). Approximately at the same anterior-posterior level as MSB, we found consistently an activation in the medial part of the upper bank of the STS, which we will label as the middle posterior upper bank STS body patch (MPUB). This patch was activated less consistently when contrasting bodies and control objects in our previous body mapping study using static images (Popivanov et al., 2012). More anteriorly, another dynamic body patch was observed in the upper bank of the STS, which we will label as the middle anterior upper bank (MAUB) body patch. This patch has not been observed consistently with static images, although a hint is present in the Popivanov et al. (2012) data.

Even more anterior, we observed a body patch complex that consisted of 3 distinct patches. The latter could be differentiated since they occupied slightly different anterior-posterior positions in the different monkeys. One patch was located at the lip of the lower bank of the STS and corresponds with a patch activated by static images (Bao et al., 2020; Popivanov et al., 2012) which we have labeled before as the anterior STS body patch (ASB (Kumar et al., 2017)). Dorsal to it, in the upper bank of the STS, we observed a distinct body patch, which we will label as the anterior lateral upper bank STS (ALUB) body patch. Hints of it can be seen in the (Popivanov et al., 2012) body mapping study using static images. This patch appears to be part of the STS that responds to both auditory and visual stimuli (Baylis et al., 1987). A third body patch of the anterior body patch complex was present medially in the upper bank of the STS. We will label this patch as the anterior medial upper bank body patch (AMUB). Weak body activations can be seen at this location in the group analysis of the (Popivanov et al., 2012) static images study. More anterior and ventral to ASB, we observed the previously reported anterior ventral body patch (AVB) which is activated also with static images (Bao et al., 2020; Premereur et al., 2016). This patch was absent in most individual flat maps (**Fig. S5**), probably because of its small size. However, each monkey showed it, as demonstrated with coronal sections in **Fig. S6**. In each monkey, but at different dorsal-ventral locations, we observed body patches in the temporal pole, which we will label as temporal pole body patches (TPB). Although we focus in this paper on temporal cortical body patches, we note that body activations were also observed in the lateral intraparietal sulcus and in the frontal cortex (**Fig. S7**) and most are clearly present on the flat maps (**Fig. S5**).

3.2. Posterior body patch activations to scrambled displays

Because we observed in experiments 1 and 2 no body-selective activations in visual cortical areas V1, V2, V3, and V4 (except a small one for monkey O (**Fig. S5**)), it is highly unlikely that the activations revealed by our displays in the temporal visual areas result from low-level, spatio-temporal differences among the stimuli of the different categories. To assess directly the contribution of local spatio-temporal features to the body activations, we presented mosaic-scrambled versions of the movies to the monkeys, interleaved in blocks with the original movies in experiment 1 (3 subjects). Using the inclusive masking of the contrasts scrambled bodies – scrambled faces and scrambled bodies – scrambled

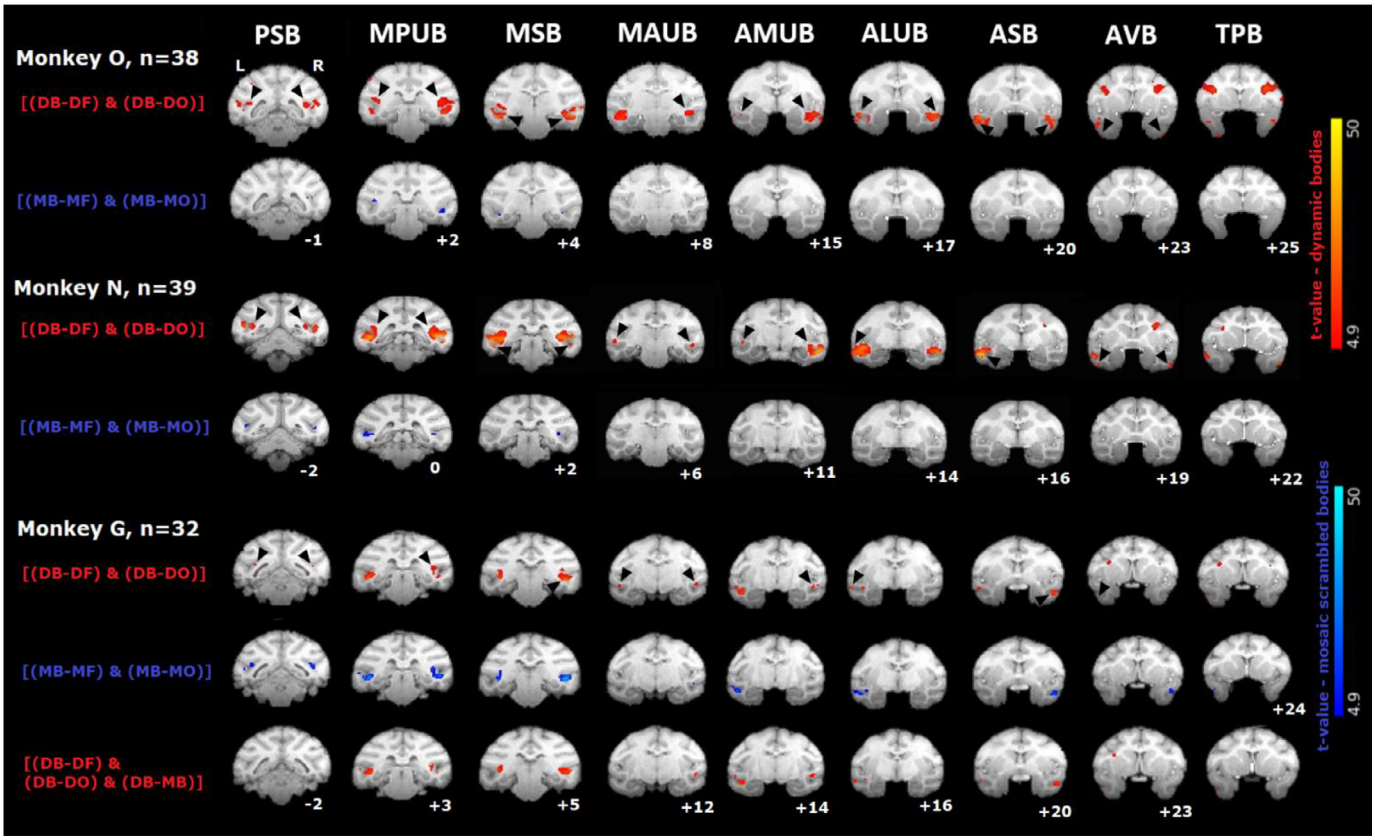

Fig. 2. Dynamic body patches and activations to mosaic-scrambled bodies (Experiment 1). Representative coronal sections of 3 monkeys (O, N and G). For each monkey, the top row of each monkey shows the body patches (red heatmap; see upper color bar) and the black arrows indicate the body patch labeled at the top of the column. The contrasts were dynamic body (DB) – dynamic object (DO) inclusively masked with dynamic body – dynamic face (DF). The second row of each monkey shows in blue (see bottom color bar) the activation to the mosaic-scrambled bodies, obtained with the contrasts mosaic-scrambled body (MB) – mosaic-scrambled object (MO) inclusively masked with mosaic-scrambled body (MB) – mosaic-scrambled face (MF). The bottom row shows the activations in monkey G when combining the contrast dynamic body (DB) – mosaic-scrambled body (MB) with the two other contrasts to define the body patch. The number of runs (n) is shown for each monkey. The distance between the coronal section and the ear canal in shown in mm (negative = more posterior) for each monkey. PSB: Posterior STS body patch; MPUB: Middle Posterior STS Upper Bank body patch; MSB: Middle STS body patch; MAUB: Middle Anterior Upper Bank STS body patch; AMUB: Anterior Medial Upper Bank body patch; ALUB: Anterior Lateral Upper Bank body patch; ASB: Anterior STS body patch; AVB: Anterior Ventral body patch; TPB: Temporal Pole body patch. L: left; R: right.

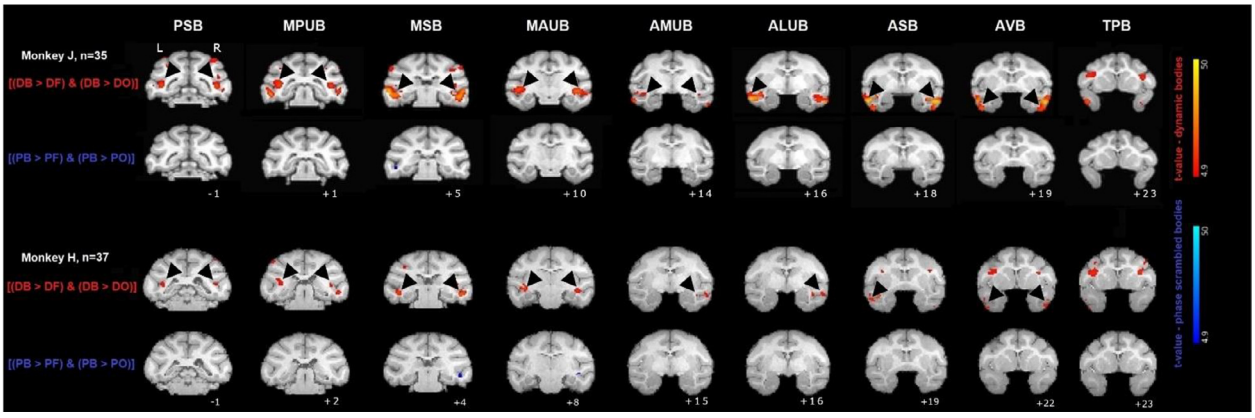

Fig. 3. Dynamic body patches and activations to phase-scrambled bodies (Experiment 2). Representative coronal sections of 2 monkeys (J and H), showing the dynamic body patches (upper row per monkey) and activations to phase-scrambled bodies (lower row per monkey). PB: phase-scrambled body; PF: phase-scrambled face; PO: phase-scrambled object. Same conventions and body patch labels as in Fig. 2. The number of runs (n) is shown for each monkey.

objects, we observed activations in body patches mapped with the original movies (Fig. 2; for flat maps see **Fig. S8**). However, there were individual differences in the activations for the mosaic-scrambled movies. In two monkeys, there were weak activations to the mosaic-scrambled bodies in only the posterior and middle STS body patches. However, in the third monkey (G) activations to the mosaic-scrambled body movies were present in body patches up to the temporal pole. In addition, activations to the mosaic-scrambled bodies were present in early visual cortex (**Fig. S8**). It is noteworthy that also the face (but not the body) patches PL, ML, AL, and AM were activated by the mosaic-scrambled faces in monkey G (**Fig. S9**). Thus, in this monkey, both body and face patch networks were activated by the mosaic-scrambled movies in a

category-specific way, i.e. face patches by mosaic-scrambled faces and body patches by the scrambled bodies. The cause of the marked difference between the third and the other two monkeys regarding the activations by the mosaic-scrambled stimuli is unclear. A psychophysical study in naïve human subjects showed that a rough animate-inanimate categorization of the mosaic-scrambled movies was possible, probably because of the presence of partial limb and eye features. Hence, one possibility is that the remaining partial body (face) features in the mosaic-scrambled images induced a body (face)-like percept in monkey G. Perhaps during the exposure to the stimuli in the fixation training preceding the scanning, the monkey learned an association between the scrambled and original movies, causing a recall of the movie during the exposure to the scrambled stimuli during scanning. Whatever the cause of the difference in the mosaic-scrambled activations among the monkeys, the body patches could still be demonstrated, even in the third monkey (except for PSB), when we included the contrast body - mosaic-scrambled body, i.e. inclusive masking of the body – face, body – object and body – mosaic-scrambled body (Fig. 2). Thus, the body-selective activations were stronger for the original compared to the mosaic-scrambled stimuli. The (albeit weaker) activations to the mosaic-scrambled stimuli in the three monkeys in posterior patches suggest that particularly these posterior body patches respond also to local features in the absence of a whole-body configuration.

Because of the activations to the mosaic-scrambled stimuli in which partial body features were still present, we employed in two other monkeys Fourier phase-scrambled versions of the movies which destroy local shape features but keep the overall motion intact (experiment 2). Using the inclusive masking of the contrasts scrambled bodies – scrambled faces and scrambled bodies – scrambled objects, activations were present only in MSB and MPUB, and these were much weaker (and only detectable in one hemisphere) than the body-selective activations for the original movies (Fig. 3; flat map: **Fig. S10**).

3.3. Stronger body-selective activations for dynamic than static displays in most body patches

To assess whether the body-selective activations revealed by the dynamic stimuli require motion, we presented static frames of the movies, interleaved in blocks with the dynamic stimuli in experiment 3. We selected 2 distinct frames per movie and the 40 images per category were presented in randomized order within a block. The exposure duration of an image was 500 ms, which is the same duration as used by Popivanov et al. (2012). By design, we choose a shorter duration of the static stimuli than the 1 s movie duration to reduce the adaptation effect that is expected to occur during a long presentation of a static stimulus (Vogels, 2016). Blocks of static images were presented randomly interleaved in the same runs as those of the dynamic movies, allowing a within-run comparison of the activations for dynamic and static stimuli.

Fig. 4 provides a direct comparison between the body-selective activations for the dynamic and static stimuli obtained in the same runs of experiment 3, employing the same contrasts (inclusive masking bodies – objects and bodies – faces) for both sorts of stimuli (experiment 3; flat maps in **Fig. S11**). In monkey H, PSB and all upper bank STS patches failed to activate for the static images whereas MSB, ASB, and AVB, the typical static body patches, were activated. In monkey J, who showed overall stronger activations than the other monkey in this experiment, PSB and upper bank body patch activations were present also for the static images, except for ALUB. However, in monkey J, the rostral upper bank activations to static images were small, not showing up on the flat map (**Fig. S11**). Across the two monkeys, only MSB, ASB, and AVB were activated for the static images, in agreement with previous imaging data from different labs. Interestingly, in monkey J, the parietal (Fig. 4) and frontal body patches were also weakly activated by static images (**Fig. S11**). Also, monkey J showed in experiment 3, but not in experiment 2, a small activation in the dorsal early visual cortex for both static and dynamic body stimuli (**Fig. S11**).

To assess the effect of dynamics for the body stimuli, we computed the contrast dynamic bodies – static bodies. This contrast showed in each monkey the expected motion-related activations in striate- and extrastriate visual cortex, including MT and FST (Fig. 5). Anterior to FST, stronger responses to the dynamic compared to the static bodies were present in most body patches, although this effect of motion was relatively weak in the anterior patches (Fig. 5). Thus, these data show that motion increased the activation to bodies in the body patches, which can explain why we consistently observed more body patches with dynamic displays than when using static images. To quantify the effect of motion and category selectivity for static versus dynamic stimuli in the body patches, we computed the percent signal change in each of the body patches that were defined in each monkey using the independent dynamic stimuli data of experiment 2 (body patch ROIs shown in **Fig. S12**) To facilitate the comparison between patches, we normalized the percent signal change for each patch and category by the mean percent signal change for the dynamic bodies. A repeated measures ANOVA showed a main effect of dynamic versus static in PSB, all middle STS patches, and ASB, while for the anterior upper bank patches and AVB motion effects were weaker and significance varied between the monkeys (Fig. 5). Upper bank STS patches tended to have a stronger effect of motion compared to lower bank STS or ventral IT patches. All patches, except the anterior upper bank STS ones and AVB showed a significant interaction of the dynamic versus static and category factors (Fig. 5). The middle upper bank STS patches, especially MAUB in both monkeys, were activated similarly by static bodies and faces, thus losing their selectivity for bodies versus faces when stimuli became static.

3.4. Similar face-selective activations for dynamic and static displays in most face patches

We defined the face patches by inclusive masking the contrasts faces – objects and faces – scrambled faces. Face patches were observed close to body patches, as observed before for static stimuli (Bao et al., 2020; Popivanov et al., 2012). We employed established terminology of Fisher and Freiwald (2015a), Hesse and Tsao (2020), Landi and Freiwald (2017) to label the face patches in monkeys J and H (experiment 2). We could identify face patches PL, ML, MF, MD, AL, AF, AD, AM, and TP with our dynamic stimuli in experiment 2 in each of the monkeys (**Fig. S13**).

Subtracting the static faces from the dynamic faces revealed activations in striate and extrastriate visual areas, including MT and FST. In both monkeys, we observed also an activation more anterior in the fundus of the STS partially overlapping face patch MF and body patch MSB (Fig. 6). However, unlike for the body patches, no motion-related activation was present in the whole-brain analysis for other face patches in both monkeys (Fig. 6). The effect of motion on the category responses was examined quantitatively by computing the percent signal changes in each of the face patch ROIs defined using the independent data of experiment 2 (Fig. 6). In monkey H, only 3 (PL, MF, and MD) out of the 9 face patches, showed a significant effect of the dynamic versus static factor, whereas in the other monkey this was true for 4 patches (ML, MF, MD, and AF; repeated measures ANOVA). However, the overall effect of motion was numerically rather small, except for MF. Two patches (PL and MF) showed a significant interaction of the dynamic versus static and category factors, and this only in one of the monkey subjects. The interaction was mainly due to a decrease in activation for the non-face categories: objects for PL and bodies and objects for MF. Note that in both monkeys the MD ROI, defined in experiment 2, failed to show clear face category selectivity, being equally activated by bodies and faces in experiment 3.

These analyses suggest that, unlike the neighboring body patches, the face patches were only little affected by motion in the displays employed in the present study. To compare the difference between the face and body patches in their sensitivity to motion for the same stimuli, we subtracted the normalized percent signal change to static stimuli

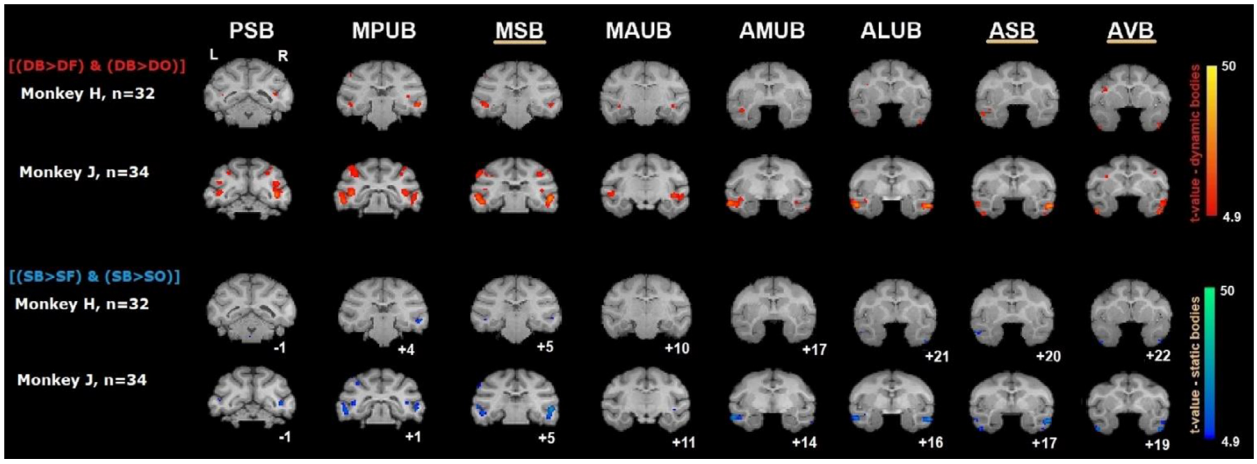

Fig. 4. Body patches obtained with dynamic and static displays (Experiment 3). Representative coronal sections of 2 monkeys (J and H), showing the body patches obtained with the dynamic displays (upper two rows) and static images (bottom two rows). SB: static body; SF: static face; SO: static object. The underlined patch labels correspond to patches that were present in both subjects for dynamic and static stimuli. Same conventions and body patch labels as in Fig. 2. The number of runs (n) is shown for each monkey.

from that to the dynamic stimuli for the body patch and face patch ROIs (Fig. 7). To reveal consistent effects for the two subjects, we pooled the data of both monkeys (for individual monkey data, see **Fig. S14**). Except for the most anterior body patch AVB, all body patches showed a significantly stronger response to dynamic compared to static bodies (i.e., a zero value outside the bootstrapped 95% confidence intervals of the mean difference). The same dynamic body displays activated only face patch MF more than static bodies. This was not merely due to bodies being a less effective stimulus in face patches, since also the more effective dynamic faces produced a (marginally significant) stronger response than static faces in only one face patch, again MF. Furthermore, the same dynamic face displays showed a stronger activation than the static snapshots in 4 of the 8 examined body patches. Hence, body patches show greater sensitivity to dynamic versus static displays of bodies and faces than face patches.

3.5. No evidence for eye-movement-driven body patch activations

Because dynamic stimuli can drive eye-movements which can induce visual stimulation that result in activations and eye movement patterns might depend on the visual category, we analyzed the eye movements of each of the monkey subjects in the three experiments. Although the average percent of fixation inside the fixation window differed numerically among the 6 stimulus conditions, these differences were small (maximally 5%) and not consistent across animal subjects in the different experiments (**Fig. S15**). Furthermore, for each subject and experiment, the mean percent of fixation was above 89% for each stimulus condition. The distributions of the eye positions were highly similar across the different stimulus conditions in the different experiments **(Fig. S16)**, except for a small tendency towards an upward bias in the eye positions for the face conditions in both monkeys of experiment 3. However, this bias was present for dynamic and static faces and thus is unlikely to have caused a difference in activation between these two types of stimuli.

The saccade frequency and mean saccade amplitude (**Fig. S17**) did not differ between the 6 stimulus conditions consistently across monkeys in each of the experiments. The largest difference in saccade frequency and amplitude was observed between the noise background fixation condition and the six stimulus conditions, probably because of the absence of an image or video in the former condition. Interestingly, there was no consistent difference across monkeys in saccade frequency, nor in saccade amplitude, between the dynamic and static stimulus conditions. The distributions of saccade direction showed idiosyncratic patterns that differed between monkeys of the same experiment, but this pattern was similar across the 6 stimulus conditions (**Fig. S18**), without across-monkey consistent effects between conditions.

To assess whether eye velocity was differentially affected by stimulus category, we plotted the distribution of instantaneous speed for horizontal and vertical eye movement directions for the different stimulus categories for each monkey and experiment (**Fig. S19**). These eye speed distributions did not differ consistently between stimulus categories across monkeys. Also, the eye speed distributions were similar between the static and dynamic stimuli in each monkey of experiment 3.

In sum, we did not find evidence for the hypothesis that putative differences in eye movement patterns amongst the stimulus categories drove the body-selective activations we observed across 5 monkeys in this set of experiments.

4. Discussion

We mapped patches in the macaque inferior temporal cortex that are activated by natural movies of acting monkeys. We observed 9 body-category selective patches, starting posteriorly in FST and ending anteriorly in the temporal pole (Fig. 8). Patches were present in and ventral to the lower bank of the STS ("ventral" patches in Fig. 8) and in its upper bank ("dorsal" patches in Fig. 8). The body patches were close to face patches in both banks of the STS and anteriorly ventral to the STS. Body patches were activated more strongly by dynamic videos than by static stimuli, but this was less the case for the neighboring face patches. This suggests that dynamics have a stronger effect on body compared to face patches.

The use of dynamic bodies revealed a larger number of body patches than those found typically when using static images. Indeed, in addition to the classical body patches in the lower bank of the STS (MSB and ASB) and one ventral to the STS (AVB), we observed several body patches in the upper bank of the STS. Upper bank STS body patches have been described before when comparing static body and object images (Popivanov et al., 2012), but were less consistent amongst monkeys. Also, unlike for dynamic stimuli, upper bank body patches mapped with static images were not consistent among the two monkey subjects in the present study. Hence, the use of dynamic stimuli provides a more robust mapping of patches involved in body processing, in particular those belonging to the upper bank STS. In agreement with the present study, it has been reported before that upper bank body patches showed similar activations to static bodies and faces (Jastorff et al., 2012). The local motion energy of the body displays was larger than

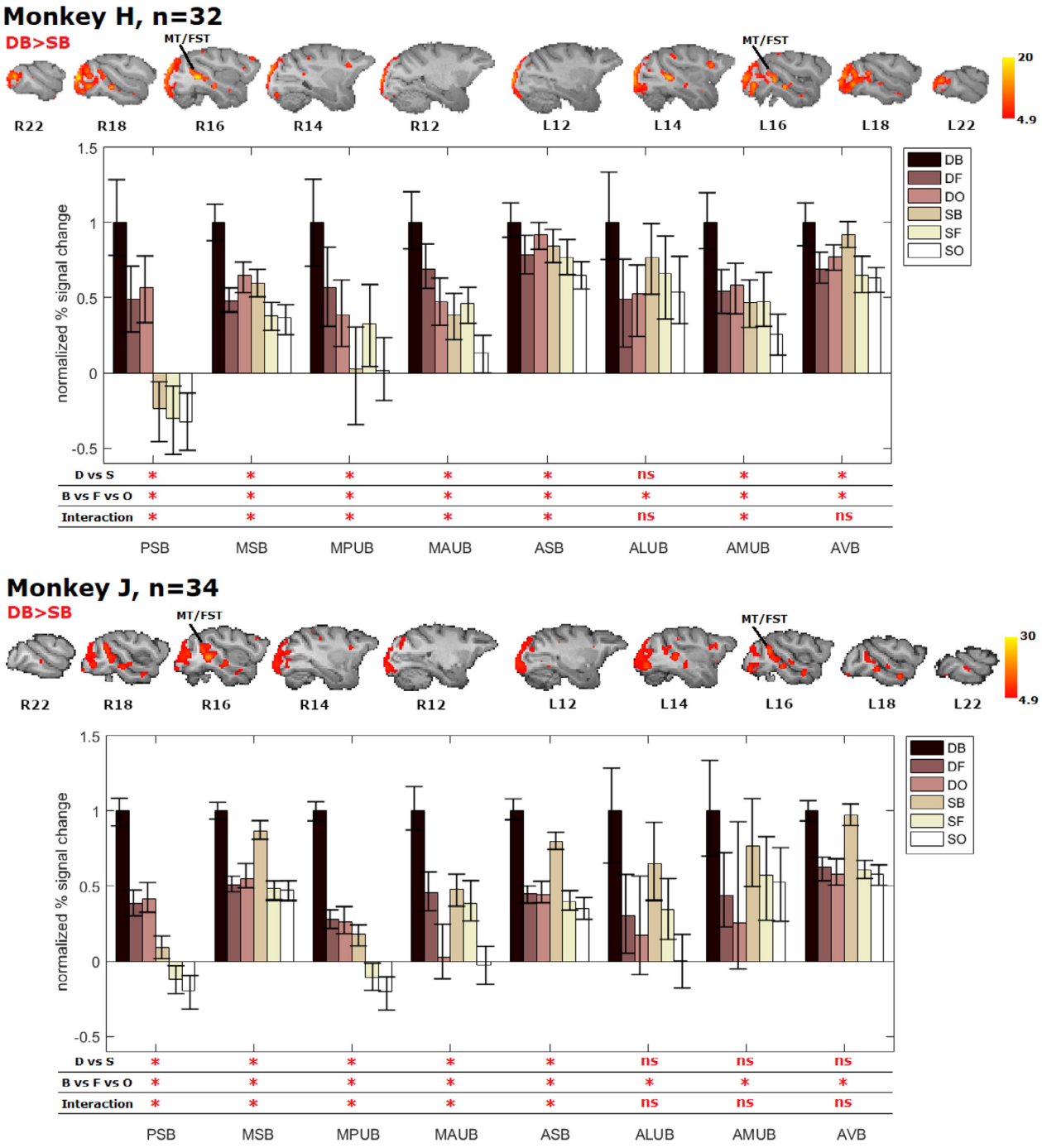

Fig. 5. Percent signal change for dynamic and static bodies, faces, and objects in body patches defined by dynamic bodies. Percent signal change for each body patch ROI, normalized by the mean percent signal change for dynamic bodies per ROI, is plotted for two monkeys separately. The percent signal changes were computed using the data of experiment 3, whereas the ROIs were defined based on experiment 2. Error bars are 95% confidence intervals obtained by bootstrapping runs. Above each panel, activations (red heatmap; see color bar) for the contrast dynamic bodies (DB) – static bodies (SB) are shown on parasagittal sections. Numbers indicate the distance from the midline for the right (R) and (L) hemispheres. The three rows below each panel show the result of an ANOVA. The first row indicates the main effect of motion (dynamic (D) versus static (S)), the second row the main effect of category (body (B), face (F), and object (O)), and the third row the interaction of the two factors. Ns: not significant (p => 0.05). The number of runs (n) are shown for each monkey. DF: dynamic face; DO: dynamic object; SF: static face; SO: static object. For the definition of body patches in these monkeys, see Fig. 3. Same body patch labels as in Fig. 3.

those of the faces and this difference in motion energy was also present for the mosaic-scrambled displays. The interaction (dynamic bodies – mosaic-scrambled bodies) – (dynamic faces – mosaic-scrambled faces) showed the same upper bank STS activations (data not shown), suggesting the stronger activations to dynamic bodies relative to faces is not merely due to local motion differences between the body and face displays. Anyhow, the upper bank patches were activated stronger by bodies compared to objects and differed from neighboring face patches. Posteriorly, we observed activations in posterior motion areas MT and FST. The FST activation was consistently observed in each of the animals and showed a strong effect of dynamics, as expected from motion areas. MT and FST are more activated by a grasping hand than

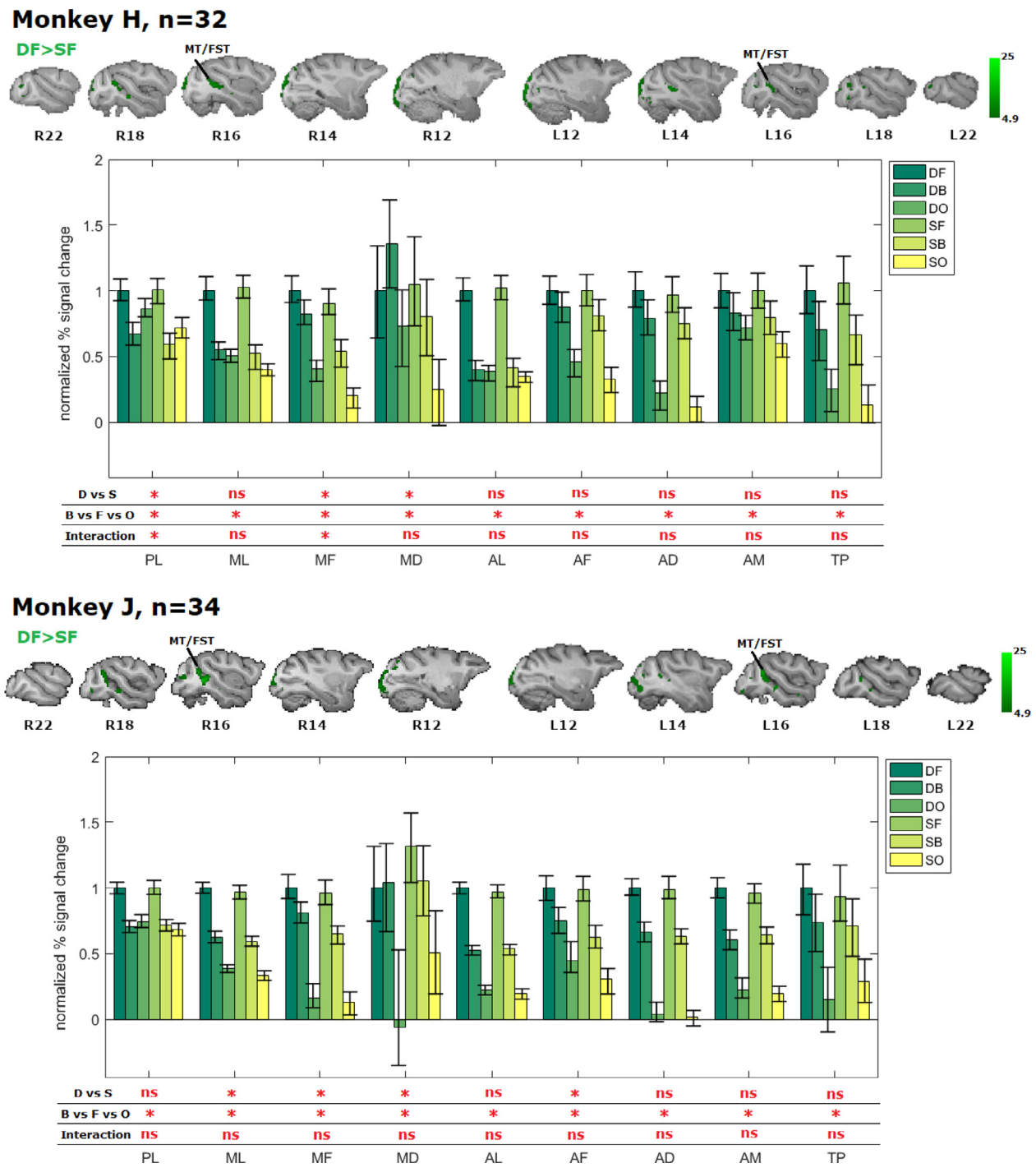

Fig. 6. Percent signal change for dynamic and static bodies, faces and objects in face patches defined by dynamic faces (Experiment 3). Percent signal change for each face patch ROI, normalized by the mean percent signal change for dynamic faces per ROI, is plotted for two monkeys separately. The percent signal changes were computed using the data of experiment 3, whereas the ROIs were defined based on experiment 2. Error bars are 95% confidence intervals obtained by bootstrapping runs. Above each panel, activations (green; see color bar) for the contrast dynamic faces (DF) – static faces (SF) are shown on parasagittal sections. Numbers indicate the distance from the midline for the right (R) and (L) hemispheres. The three rows below each panel show the result of an ANOVA. The first row indicates the main effect of motion (dynamic (D) versus static (S)), the second row the main effect of category (body (B), face (F), and object (O)), and the third row the interaction of the two factors. Ns: not significant (p => 0.05). The number of runs (n) is shown for each monkey. DB: dynamic body; DO: dynamic object; SB: static body; SO: static object. For the definition of face patches in these monkeys, see Fig. S13.

a moving object (Nelissen et al., 2006), and thus may be involved in the analysis of body kinetics. In both monkeys tested with static images, the FST/PSB activations tended to be higher for bodies compared to faces and objects (statistically significant in one monkey), which suggests that differences in motion features might not be the only factor causing the enhanced activations to the body displays in this patch.

Mapping activations to predefined natural categories such as bodies and faces do not inform us about the features that drive the category-selective responses. Recently, Bao et al. (2020) suggested the presence

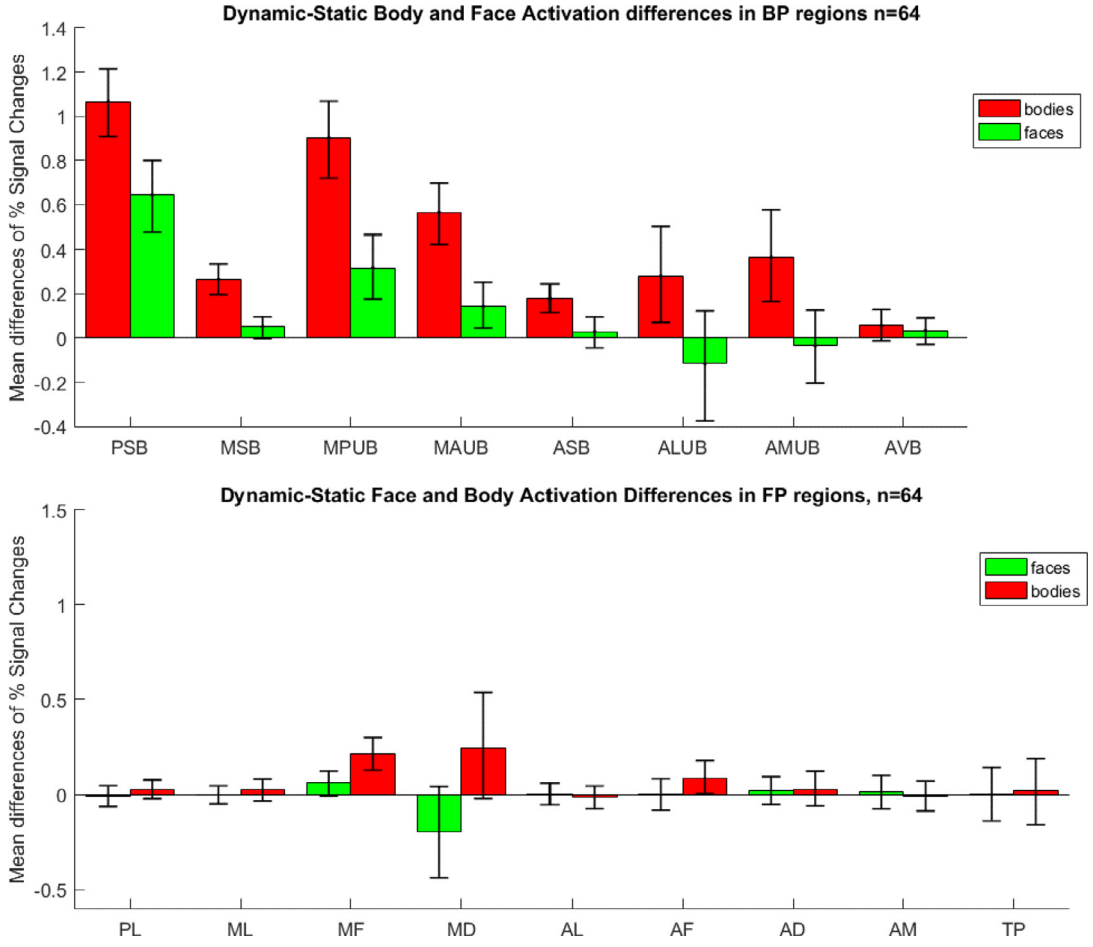

Fig. 7. Comparison of the effect of dynamics on activations for the same displays between body and face patches. Difference in normalized percent signal change between dynamic and static displays of bodies (red) and faces (green) in body patches (BP; **A**) and face patches (FP; **B**). Error bars represent 95% confidence intervals, obtained by bootstrapping differences in activation across the 64 runs (data pooled of two monkeys after equating the number of runs for each monkey; experiment 3). For each body patch ROI, we normalized the percent signal change for each category by dividing it by the percent signal change for the dynamic bodies. Likewise, for each face patch ROI, we normalized the percent signal change by dividing these by the percent signal change for the dynamic faces. To compare the effect of motion for the same stimuli between the face and body patches, we subtracted the normalized percent signal change for the static stimuli from the normalized percent signal change for the dynamic stimuli. This was done after pooling the normalized percent signal change data of the two monkeys. Same body and face patch labels as in Figs. 5 and 6, respectively.

of an object category map in the macaque IT reflecting quadrants of a space defined by the first two principal components (PC) of the activations of a large set of images in a deep convolutional neural network layer that was trained on object classification. One of the quadrants of the two-dimensional PC space was populated by images of mammal and bird bodies and mapped to MSB, ASB, and AVB. How this object space proposed for static images relates to the activations in the upper bank of the STS that we revealed here by using dynamic bodies is unclear.

Mosaic-scrambling of the dynamic displays reduced but did not abolish activations in the body patches. Our psychophysical study in human subjects showed that some features indicative of animate objects remained visible in our mosaic-scrambled movies which may have contributed to the remaining activations to the scrambled bodies. The activations to the mosaic-scrambled bodies were mainly in the posterior and middle STS body patches. Anterior body patches may show greater sensitivity to body configuration, explaining their lower activation by mosaic-scrambled displays. The latter fits the higher hierarchical stage of the anterior relative to the posterior body patches (Bao et al., 2020; Kumar et al., 2017). Our data suggest that this holds for both ventral and dorsal bank body patches. Interestingly, Fourier-phase scrambling, which abolishes local form features, still produced activations in the middle STS patches. The Fourier-phase scrambling kept the global motion of the original displays. The weak activations to these scrambled displays may suggest that the middle STS body patches are sensitive to motion patterns that are inherent to bodily movements.

Overall, we observed a stronger activation for the dynamic than for static stimuli in several body patches, and a larger number of patches were found for dynamic compared to static stimuli. The stronger activation for dynamic compared to static stimuli tended to be more pronounced in the middle STS than in most anterior body patches. This agrees with a human fMRI study that found stronger activation to dynamic images in the posterior body area EBA but only a nonsignificant tendency in the more anterior FBA (Pitcher et al., 2019).

Several factors can in principle cause a difference in activation between dynamic and static stimuli. First, dynamic displays can include a greater variation in postures and body orientation than the static images, thus producing a higher activation to the dynamic stimuli in posterior orientation- and posture-selective body patch neurons (Bao et al., 2020; Kumar et al., 2017). However, we compared the dynamic displays with 40 highly different static images selected from those displays. Thus, it is unlikely that this factor can explain the difference in activation between dynamic and static displays, especially for some of the upper bank STS patches which lost their selectivity for bodies versus faces

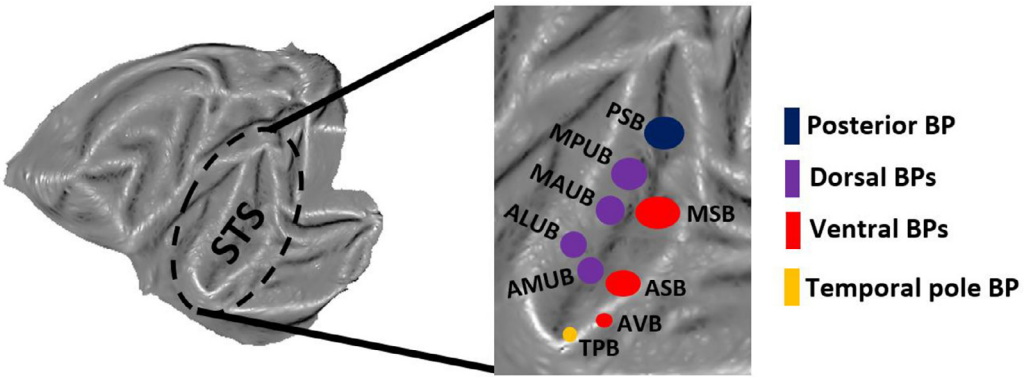

Fig. 8. Schematic of temporal cortical body patches revealed by dynamic stimuli. Three body patches ("ventral"), which are activated also by static images, are located in the middle and rostral lower bank of the STS and ventral to the lower bank of the STS. Four patches are in the middle and rostral upper bank of the STS ("dorsal"). Body-selective activations were also present more posteriorly in the STS (PSB) and the temporal pole (TPB).

when stimuli became static. Previous fMRI studies addressed the contribution of this image variation factor for faces, by comparing sequences with natural and randomly ordered frames (Fisher and Freiwald, 2015a; Schultz et al., 2013). However, a higher activation to the natural compared to the random sequences can result from temporal summation for the similar successive frames in the case of the natural sequences and a reduction in response in the case of the randomly ordered sequences because of forward and backward masking, which renders such an activation difference difficult to interpret.

Second, comparing the fMRI activations between static and dynamic displays is not straightforward because of the difficulty in estimating the effect of stimulus duration on the activation for the static images. A block of long-duration stimuli is expected to produce a weaker response than an equally long block of a higher number of short-duration stimuli because of the strong transient response of visual cortical neurons, even in IT. The relatively weaker sustained response phase, as a result of spike-rate adaptation, will dominate the response to long-duration stimuli, thus producing a smaller response than a succession of shorter duration stimuli which will produce relatively more transient responses. Indeed, a succession of brief stimuli produces a stronger BOLD response than longer stimuli in human ventral visual stream areas (Stigliani et al., 2015). To reduce the contribution of this factor we employed a duration of 500 ms for the static stimuli, which is shorter than the duration of the dynamic stimuli. The choice of 500 ms duration was also based on previous monkey fMRI mapping of body patches with static stimuli (Popivanov et al., 2012), and also allowed presenting a wide range of static images (see above). Because of the short static stimulus duration we employed in the current study, we believe it is unlikely that the stimulus duration factor can explain the greater response to the dynamic bodies in our study.

Third, the greater activation for the dynamic versus static images can result from the smooth motion present in the dynamic images. The static images in our experiment were presented back-to-back, without an interstimulus interval, which produced also motion when going from one static image to the next. However, this transient, random motion in the static stimulus blocks differs from the smooth, natural motion of the dynamic displays. Our Fourier-phase scrambled displays contained the same overall motion as the dynamic body displays but produced a much lower body-selective activation than the latter. This suggests that the responses to the dynamic displays are not driven only by global motion per se but requires also particular configurations of motion or shape features. The robust body-selective activation of ventral STS body patches to the static stimuli supports a strong contribution of spatial features to their responses. Upper bank STS patches were activated more consistently by the dynamic displays than by the static stimuli. This agrees with single-unit studies that suggest a stronger contribution of motion in the upper bank of the macaque STS compared to the lower bank of the STS and the inferior temporal cortical convexity (Baylis et al., 1987; Vangeneugden et al., 2011, 2009).

In contrast to the neighboring body patches, dynamic and static stimuli showed comparable activations in most face patches. Across monkeys, we found only in the fundus STS face patch MF an enhanced activation for dynamic relative to static stimuli. Previous monkey fMRI studies showed more profound differences between dynamic and static stimuli in face patches (Fisher and Freiwald, 2015a; Polosecki et al., 2013; Zhang et al., 2020). In two of these studies, this could have resulted from the 2 s (Zhang et al., 2020) or the up to 2.5 s exposure duration (Polosecki et al., 2013) of the static stimuli, which included only a single frame of each movie. The combination of the smaller variation of faces and spike-rate adaptation because of the long exposure duration for the static stimuli might have caused the lower response for these stimuli compared to their dynamic faces. In the third study (Fisher and Freiwald, 2015a), the static blocks consisted of 3 secs long sequences of 6 faces of the same identity, each shown for 500 ms, which also may have caused adaptation. Notably, the overall pattern of the results differed among the previous studies: one study observed a significant interaction between the factors dynamic-static and category in anterior face patches AM and AL (Polosecki et al., 2013), while another study observed such interaction in only the dorsal bank STS patch AF (Zhang et al., 2020). The third study stressed the importance of motion for the dorsal bank STS face patch MD (Fisher and Freiwald, 2015a). Somewhat surprisingly, this patch was not demonstrated in the Zhang et al. study. However, a consistent observation in these studies and our study, is the presence of face-selective activations with static faces in the dynamic face patches, indicating that motion is not required to activate the face patches. This is corroborated by single-unit studies that showed face-selective responses using static images in the upper bank/fundus STS patches MF (Freiwald and Tsao, 2010), AF (Koyano et al., 2021), and MD (Yang and Freiwald, 2021). Several human fMRI studies showed weak if any, enhanced activation to dynamic compared to static stimuli in ventral face-category selective areas (OFA; FFA), but a robust effect of motion in the STS (Pitcher et al., 2011, 2019; Polosecki et al., 2013; Zhang et al., 2020). This fits our results showing little if any enhanced activation for dynamic stimuli in lower bank STS face patches and enhanced activation for the dynamic faces (and bodies) in upper bank STS face patch MF, since macaque lower bank and upper bank STS face patches are likely homologs of human ventral and STS face category-selective areas (Pitcher and Ungerleider, 2021).

With identical stimulus displays, we observed a greater effect of the dynamic versus static factor in body patches compared to the neighboring face patches. This suggests that dynamics contribute more to body than face processing. The dynamics of faces relate mainly to emotional expression and vocalizations (non-rigid motion) and changes in 3D orientation (rigid motion) (O'Toole et al., 2002). Our face stimulus set con-

tained both types of dynamics. Face patch neurons are selective for the 3D orientation of static stimuli (Freiwald and Tsao, 2010; Yang and Freiwald, 2021), and at least upper bank STS neurons are selective for the emotional expression of static faces (Yang and Freiwald, 2021). Thus, static face features appear to be sufficient to drive many face-selective neurons.

We also observed body-selective activations in the frontal and parietal cortex. These appear to overlap regions previously revealed in fMRI studies of action observation in macaques (e.g. Cui et al. 2022, Fiave et al. 2018). Interestingly, the frontal and parietal activations were also weakly present for static bodies in one monkey, the same one that also showed upper bank activations to static bodies. It is highly unlikely that the frontal and parietal body activations reflect differences in eye movements between stimulus conditions because we did not find evidence for a consistent difference between eye movement metrics among the dynamic stimulus categories.

Psychophysical studies in humans show that bodies in motion aid person and expression recognition, especially when face information is poor, e.g. when bodies are seen from a distance (Yovel and O'Toole, 2016). Recognition of actions, including bodily emotional expressions, can benefit from information about the sequence of postures and motion features as is present in dynamic displays of actions, but absent in static displays. Which factors contribute to the enhanced activation to dynamic body displays is a topic of further inquiry. The present fMRI experiments identified a set of patches in the macaque temporal cortex that are activated by dynamic body stimuli, laying the foundation for future single-unit recording studies to reveal the spatiotemporal features the neurons of these patches encode.

Data and code availability

Preprocessed fMRI data and the body stimuli are available at https://data.mendeley.com/datasets/n8v6p7gfjg/1. Codes are available upon request.

Declaration of Competing Interest

The authors declare no competing interests.

Credit authorship contribution statement

A. Bognár: Conceptualization, Methodology, Writing – review & editing. **R. Raman:** Conceptualization, Methodology, Investigation, Writing – review & editing. **N. Taubert:** Methodology. **Y. Zafirova:** Methodology. **B. Li:** Conceptualization, Investigation. **M. Giese:** Conceptualization, Methodology, Writing – review & editing, Funding acquisition. **B. De Gelder:** Conceptualization, Writing – review & editing, Funding acquisition. **R. Vogels:** Conceptualization, Methodology, Writing – review & editing, Funding acquisition, Supervision.

Data availability

Preprocessed fMRI data and the body stimuli are available at https://data.mendeley.com/datasets/n8v6p7gfjg/1. Codes are available upon request.

Acknowledgments

The authors thank A. Coeman, C. Fransen, P. Kayenbergh, I. Puttemans, A. Hermans, G. Meulemans, W. Depuydt, C. Ulens, S. Verstraeten, and M. De Paep for technical and administrative support. We are indebted to Dr. Qi Zhu, Dr. K. Nelissen and Dr. W. Vanduffel for methodological advice and use of the facial expression video material. We thank Dr. D. Cui for invaluable assistance in preparing the flat maps. This research was supported by Fonds voor Wetenschappelijk Onderzoek (FWO) Vlaanderen (G0E0220N), KU Leuven grant C14/21/111 and by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 856495).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2023.119907.

References

Bao, P., She, L., McGill, M., Tsao, D.Y., 2020. A map of object space in primate inferotemporal cortex. Nature 583, 103–108.
Bao, P., Tsao, D.Y., 2018. Representation of multiple objects in macaque category-selective areas. Nat. Commun. 9, 1774.
Baylis, G.C., Rolls, E.T., Leonard, C.M., 1987. Functional subdivisions of the temporal lobe neocortex. J. Neurosci. 7, 330–342.
Bell, A.H., Hadj-Bouziane, F., Frihauf, J.B., Tootell, R.B., Ungerleider, L.G., 2009. Object representations in the temporal cortex of monkeys and humans as revealed by functional magnetic resonance imaging. J. Neurophysiol. 101, 688–700.
Bell, A.H., Malecek, N.J., Morin, E.L., Hadj-Bouziane, F., Tootell, R.B., Ungerleider, L.G., 2011. Relationship between functional magnetic resonance imaging-identified regions and neuronal category selectivity. J. Neurosci. 31, 12229–12240.
Cox, D., Meyers, E., Sinha, P., 2004. Contextually evoked object-specific responses in human visual cortex. Science 304, 115–117.
Cui, D., Sypre, L., Vissers, M., Sharma, S., Vogels, R., Nelissen, K., 2022. Categorization learning induced changes in action representations in the macaque STS. Neuroimage 265, 119780.
de Gelder, B., de Borst, A.W., Watson, R., 2015. The perception of emotion in body expressions. Wiley Interdiscip. Rev. Cogn. Sci. 6, 149–158.
Ekstrom, L.B., Roelfsema, P.R., Arsenault, J.T., Bonmassar, G., Vanduffel, W., 2008. Bottom-up dependent gating of frontal signals in early visual cortex. Science 321, 414–417.
Engbert, R., Kliegl, R., 2003. Microsaccades uncover the orientation of covert attention. Vis. Res. 43, 1035–1045.
Engbert, R., Mergenthaler, K., 2006. Microsaccades are triggered by low retinal image slip. Proc. Natl. Acad. Sci. USA 103, 7192–7197.
Fiave, P.A., Sharma, S., Jastorff, J., Nelissen, K., 2018. Investigating common coding of observed and executed actions in the monkey brain using cross-modal multi-variate fMRI classification. Neuroimage 178, 306–317.
Fisher, C., Freiwald, W.A., 2015a. Contrasting specializations for facial motion within the macaque face-processing system. Curr. Biol. 25, 261–266.
Fisher, C., Freiwald, W.A., 2015b. Whole-agent selectivity within the macaque face-processing system. Proc. Natl. Acad. Sci. USA 112, 14717–14722.
Freiwald, W.A., Tsao, D.Y., 2010. Functional compartmentalization and viewpoint generalization within the macaque face-processing system. Science 330, 845–851.
Giese, M.A., Poggio, T., 2003. Neural mechanisms for the recognition of biological movements. Nat. Rev. Neurosci. 4, 179–192.
Grezes, J., Pichon, S., de Gelder, B., 2007. Perceiving fear in dynamic body expressions. Neuroimage 35, 959–967.
Hesse, J.K., Tsao, D.Y., 2020. The macaque face patch system: a turtle's underbelly for the brain. Nat. Rev. Neurosci. 21, 695–716.
Jastorff, J., Popivanov, I.D., Vogels, R., Vanduffel, W., Orban, G.A., 2012. Integration of shape and motion cues in biological monkey STS. Neuroimage 60, 911–921.
Kaliukhovich, D.A., Vogels, R., 2011. Stimulus repetition probability does not affect repetition suppression in macaque inferior temporal cortex. Cereb. Cortex 21, 1547–1558.
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. J. Neurosci. 29, 7031–7039.
Koyano, K.W., Jones, A.P., McMahon, D.B.T., Waidmann, E.N., Russ, B.E., Leopold, D.A., 2021. Dynamic suppression of average facial structure shapes neural tuning in three macaque face patches. Curr. Biol. 31, 1–12 e15.
Kumar, S., Popivanov, I.D., Vogels, R., 2017. Transformation of Visual Representations Across Ventral Stream Body-selective Patches. Cereb. Cortex 29, 215–229.
Lafer-Sousa, R., Conway, B.R., 2013. Parallel, multi-stage processing of colors, faces and shapes in macaque inferior temporal cortex. Nat. Neurosci. 16, 1870–1878.
Landi, S.M., Freiwald, W.A., 2017. Two areas for familiar face recognition in the primate brain. Science 357, 591–595.
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., Mandeville, J.B., 2002. Repeated fMRI using iron oxide contrast agent in awake, behaving macaques at 3 Tesla. Neuroimage 16, 283–294.
Nelissen, K., Vanduffel, W., Orban, G.A., 2006. Charting the lower superior temporal region, a new motion-sensitive region in monkey superior temporal sulcus. J. Neurosci. 26, 5929–5947.
O'Toole, A.J., Roark, D.A., Abdi, H., 2002. Recognizing moving faces: a psychological and neural synthesis. Trends Cogn. Sci. 6, 261–266.
Peelen, M.V., Downing, P.E., 2007. The neural basis of visual body perception. Nat. Rev. Neurosci. 8, 636–648.
Pinsk, M.A., Arcaro, M., Weiner, K.S., Kalkus, J.F., Inati, S.J., Gross, C.G., Kastner, S., 2009. Neural representations of faces and body parts in macaque and human cortex: a comparative FMRI study. J. Neurophysiol. 101, 2581–2600.

Pinsk, M.A., DeSimone, K., Moore, T., Gross, C.G., Kastner, S., 2005. Representations of faces and body parts in macaque temporal cortex: a functional MRI study. Proc. Natl. Acad. Sci. USA 102, 6996–7001.
Pitcher, D., Dilks, D.D., Saxe, R.R., Triantafyllou, C., Kanwisher, N., 2011. Differential selectivity for dynamic versus static information in face-selective cortical regions. Neuroimage 56, 2356–2363.
Pitcher, D., Ianni, G., Ungerleider, L.G., 2019. A functional dissociation of face-, body- and scene-selective brain areas based on their response to moving and static stimuli. Sci. Rep. 9, 8242.
Pitcher, D., Ungerleider, L.G., 2021. Evidence for a third visual pathway specialized for social perception. Trends Cogn. Sci. 25, 100–110.
Polosecki, P., Moeller, S., Schweers, N., Romanski, L.M., Tsao, D.Y., Freiwald, W.A., 2013. Faces in motion: selectivity of macaque and human face processing areas for dynamic stimuli. J. Neurosci. 33, 11768–11773.
Popivanov, I.D., Jastorff, J., Vanduffel, W., Vogels, R., 2012. Stimulus representations in body-selective regions of the macaque cortex assessed with event-related fMRI. Neuroimage 63, 723–741.
Popivanov, I.D., Jastorff, J., Vanduffel, W., Vogels, R., 2014. Heterogeneous single-unit selectivity in an fMRI-defined body-selective patch. J. Neurosci. 34, 95–111.
Premereur, E., Taubert, J., Janssen, P., Vogels, R., Vanduffel, W., 2016. Effective connectivity reveals largely independent parallel networks of face and body patches. Curr. Biol. 26, 3269–3279.
Schultz, J., Brockhaus, M., Bulthoff, H.H., Pilz, K.S., 2013. What the human brain likes about facial motion. Cereb. Cortex 23, 1167–1178.
Sliwa, J., Freiwald, W.A., 2017. A dedicated network for social interaction processing in the primate brain. Science 356, 745–749.
Stigliani, A., Weiner, K.S., Grill-Spector, K., 2015. Temporal processing capacity in high-level visual cortex is domain specific. J. Neurosci. 35, 12412–12424.
Tsao, D.Y., Freiwald, W.A., Knutsen, T.A., Mandeville, J.B., Tootell, R.B., 2003. Faces and objects in macaque cerebral cortex. Nat. Neurosci. 6, 989–995.
Vanduffel, W., Fize, D., Mandeville, J.B., Nelissen, K., Van Hecke, P., Rosen, B.R., Tootell, R.B.H., Orban, G.A., 2001. Visual motion processing investigated using contrast agent-enhanced fMRI in awake behaving monkeys. Neuron 32, 565–577.
Vangeneugden, J., De Maziere, P.A., Van Hulle, M.M., Jaeggli, T., Van, G.L., Vogels, R., 2011. Distinct mechanisms for coding of visual actions in macaque temporal cortex. J. Neurosci. 31, 385–401.
Vangeneugden, J., Pollick, F., Vogels, R., 2009. Functional differentiation of macaque visual temporal cortical neurons using a parametric action space. Cereb. Cortex 19, 593–611.
Vergnieux, V., Vogels, R., 2020. Statistical learning signals for complex visual images in macaque early visual cortex. Front. Neurosci. 14, 789.
Vogels, R., 2016. Sources of adaptation of inferior temporal cortical responses. Cortex 80, 185–195.
Vogels, R., 2022. More than the face: representations of bodies in the inferior temporal cortex. Annu. Rev. Vis. Sci. 8, 383–405.
Yang, Z., Freiwald, W.A., 2021. Joint encoding of facial identity, orientation, gaze, and expression in the middle dorsal face area. Proc. Natl. Acad. Sci. USA 118, 1–11.
Yovel, G., O'Toole, A.J., 2016. Recognizing People in Motion. Trends Cogn. Sci. 20, 383–395.
Zhang, H., Japee, S., Stacy, A., Flessert, M., Ungerleider, L.G., 2020. Anterior superior temporal sulcus is specialized for non-rigid facial motion in both monkeys and humans. Neuroimage 218, 116878.
Zhu, Q., Nelissen, K., Van den Stock, J., De Winter, F.L., Pauwels, K., de Gelder, B., Vanduffel, W., Vandenbulcke, M., 2013. Dissimilar processing of emotional facial expressions in human and monkey temporal cortex. Neuroimage 66, 402–411.