Tolerance of Macaque Middle STS Body Patch Neurons to Shape-preserving Stimulus Transformations

Ivo D. Popivanov¹, Jan Jastorff¹, Wim Vanduffel^{1,2,3}, and Rufin Vogels¹

Abstract

■ Functional imaging studies in human and nonhuman primates have demonstrated regions in the brain that show category selectivity for faces or (headless) bodies. Recent fMRI-guided single unit studies of the macaque face category-selective regions have increased our understanding of the response properties of single neurons in these face patches. However, much less is known about the response properties of neurons in the fMRI-defined body category-selective regions (“body patches”). Recently, we reported that the majority of single neurons in one fMRI-defined body patch, the mid-STS body patch, responded more strongly to bodies compared with other objects. Here we assessed the tolerance of these neurons’ responses and stimulus preference for shape-preserving image transformations. After mapping the receptive field of the single neurons, we found that

their stimulus preference showed a high degree of tolerance for changes in the position and size of the stimulus. However, their response strongly depended on the in-plane orientation of a body. The selectivity of most neurons was, to a large degree, preserved when silhouettes were presented instead of the original textured and shaded images, suggesting that mainly shape-based features are driving these neurons. In a human psychophysical study, we showed that the information present in silhouettes is largely sufficient for body versus nonbody categorization. These data suggest that mid-STS body patch neurons respond predominantly to oriented shape features that are prevalent in images of bodies. Their responses can inform position- and retinal size-invariant body categorization and discrimination based on shape. ■

INTRODUCTION

fMRI studies in humans and monkeys have identified category-selective areas that are activated more strongly by images of faces and bodies compared with other categories (Bell, Hady-Bouziane, Frihauf, Tootell, & Ungerleider, 2009; Pinsk et al., 2009; Pinsk, DeSimone, Moore, Gross, & Kastner, 2005; Tsao, Freiwald, Knutsen, Mandeville, & Tootell, 2003; Downing, Jiang, Shuman, & Kanwisher, 2001; Epstein & Kanwisher, 1998; Kanwisher, McDermott, & Chun, 1997). Single-unit studies in the fMRI-defined patches that are activated more strongly to faces compared with other body categories showed a large proportion of face-selective neurons in these face patches (Tsao, Freiwald, Tootell, & Livingstone, 2006) and started to characterize the face-related features these neurons respond to (Ohayon, Freiwald, & Tsao, 2012; Freiwald & Tsao, 2010; Freiwald, Tsao, & Livingstone, 2009). Much less is known about the functional characteristics of the neurons in the body category-selective regions. Recently, we reported that single neurons in the macaque mid-STS body patch, a posterior body patch in the temporal cortex (Popivanov, Jastorff, Vanduffel, & Vogels, 2012), responded on average more strongly to images of bodies (human

bodies, monkey bodies, four-legged mammals, and birds) compared with other objects, including faces (Popivanov, Jastorff, Vanduffel, & Vogels, 2014). However, individual neurons showed considerable heterogeneity in their stimulus selectivity, with some single neurons responding even more strongly to nonbody images. Moreover, the majority of neurons displayed a marked selectivity for different bodies. Despite this heterogeneity at the single-unit level, the population of mid-STS body patch neurons successfully classified bodies versus nonbodies.

Here we examine for the first time fundamental functional properties of mid-STS body patch neurons, which has implications with respect to the signals these neurons provide to other regions involved in body categorization or discrimination. Given the importance of invariance to object-preserving transformations for object and, thus, also body recognition (Dicarlo, Zoccolan, & Rust, 2012), it is essential to know to what degree single neurons in the body patch tolerate changes in retinal position, size, and rotation. The first property we examined is arguably the most fundamental one: the receptive field (RF) of the single neurons, which constrains the position invariance of their responses (Goris & Op de Beeck, 2009). Next, we examined the position and size tolerance of the stimulus preference of the body patch single units. On the basis of their location along the posterior–anterior axis, one would expect that these neurons would show

¹KU Leuven, ²MGH Martinos Centre, Charlestown, MA, ³Harvard Medical School

considerable tolerance for these stimulus transformations. In addition, and extending our previous work (Popivanov et al., 2014), we assessed their tolerance for in-plane rotations of the body images.

Finally and importantly, we examined whether a transformation that preserves only shape but not shading or texture cues affects the response and the selectivity of these neurons. We asked whether the neuron's responses and selectivity are only driven by shape features or, instead, whether "material properties" (fur, skin, etc.) and/or shading, which can provide depth information and assist segmentation of internal features, are critical. Thus, we compared the responses of the neurons for the original shaded and textured images to reduced stimuli that consisted of either the object silhouette or its outline.

In summary, we examined the effect of four shape-preserving transformations on the responses of single mid-STS body patch neurons: translation, scale, rotation, and silhouette/outline. This would provide us information about the degree of tolerance of those neurons and would shed more light on the neural mechanisms of body recognition.

METHODS

Single Unit Study

Subjects

The two male rhesus monkeys (*Macaca mulatta*) were identical to the subjects of Popivanov et al. (2014). They were implanted with a headpost and a recording chamber targeting the mid-STS. Animal care and experimental procedures complied with the national and European laws, and the study was approved by the ethical committee of KU Leuven.

Recordings

Standard single-unit recordings were performed with epoxylite-insulated tungsten microelectrodes (FHC; in situ measured impedance between 1 and 1.7 M Ω) using techniques as described previously (Sawamura, Orban, & Vogels, 2006). Briefly, the electrode was lowered with a Narishige microdrive into the brain using a guide tube that was fixed in a standard Crist grid positioned within the recording chamber. After amplification and filtering between 540 Hz and 6 kHz, single units were isolated online using a custom amplitude- and time-based discriminator.

The position of one eye was continuously tracked by means of an infrared video-based tracking system (SR Research EyeLink, Ottawa; sampling rate 1 kHz). Stimuli were presented on a CRT display (Philips Brilliance 202 P4, Amsterdam; 1024 $\times$ 768 screen resolution; 75 Hz vertical refresh rate) at a distance of 57 cm from the monkey's eyes. The onset and offset of the stimulus was signaled by means

of a photodiode detecting luminance changes of a square in the corner of the display that was invisible to the animal. A digital signal processing-based computer system controlled stimulus presentation, event timing, and juice delivery while sampling the photodiode signal, eye positions, spikes, and behavioral events. Time stamps of the spikes, eye positions, stimulus, and behavioral events were stored for offline analyses.

We recorded from the left mid-STS body patch, as defined by fMRI in the same subjects by contrasting images of headless monkey bodies and control objects (for details, see Popivanov et al., 2014). The recording locations inside the body patch were identical to those of Popivanov et al. (2014).

Stimuli

Responsive neurons were searched with 100 stimuli that included 10 classes of achromatic images—monkey and human bodies (excluding the head), monkey and human faces, four-legged mammals (with head), birds (with head), fruits/vegetables, body-like sculptures, and two classes of manmade objects. These stimuli were identical to those of the main test of Popivanov et al. (2014), and further details can be found in that paper. Low-level image characteristics, such as mean luminance, mean contrast, and aspect ratio, were equated across stimulus classes. The difference between the mean aspect ratio of the monkey and human bodies was controlled for by using two classes of manmade objects—one matching the aspect ratio of the monkey bodies (objectsM) and another one matching the aspect ratio of the human bodies (objectsH). The images were resized so that the average area per class was matched across all classes, except for the objectsH and human bodies, while allowing for some variation in area (range = 3.7° to 6.7° [square root of the area]) among the exemplars in each class. The mean vertical and horizontal extent of the images was 8.3° and 6.7°, respectively.

The stimuli were also rendered with a size of 2°, 4°, and 8°, measured as the maximum of their vertical and horizontal extent. To test the effect of in-plane orientation, we rotated the 4° objects in steps of 45° around their center of mass. To determine whether shape information was sufficient to drive the neurons, we created a silhouette and an outline version of each original area-equated image. For the silhouette version, all the pixels of the object were black, whereas for the outline version only the outer contour was presented as a black line (thickness = 0.1°). The images were embedded into pink noise backgrounds (height $\times$ width: 30° $\times$ 40°). In the case of the outline versions, the noise background was present inside and outside the contour. Each image was presented on top of nine different backgrounds that varied randomly across stimulus presentations, except when mapping the RF where the background was fixed during a test. The stimuli were gamma-corrected.

To relate the single-unit responses to low-level properties of the images, we computed for each original, outline, and silhouette versions of an image its mean luminance, contrast (RMS normalized by the mean luminance), and 2-D power spectrum (computed with the Matlab [The MathWorks, Natick, MA] *fft2* function). The power was averaged across spatial frequencies (up to the Nyquist frequency and excluding DC) and orientations. In addition, we averaged the power in a “horizontal” and “vertical” band, which had bandwidths of 90° (e.g., the horizontal band contained power in a band of $90 \pm 45^\circ$). These bands could be further subdivided in two spatial frequency bands, averaging the power above or below 3 cycles/deg. These image properties were computed for each combination of stimulus and pink noise background and then averaged across the nine stimulus–background combinations.

Tests

In all tests, stimuli were presented for 200 msec each with an ISI of approximately 400 msec during passive fixation (fixation window size = $2^\circ \times 2^\circ$). The pink noise background was present throughout the task and, except when mapping the RF, refreshed simultaneously with stimulus onset. Fixation was required in a period from 100 msec prestimulus to 200 msec poststimulus. Trials were not analyzed further when fixation was aborted during a stimulus presentation. Juice rewards were given with decreasing intervals (2000–1350 msec) as long as the monkeys maintained fixation.

Search test. Neurons were tested with pseudorandom, interleaved presentations of the 100 search images (see Popivanov et al., 2014, for more details). The mean number of unaborted presentations per stimulus was 8.1, averaged across neurons. On the basis of this test, images were selected for the subsequent tests.

RF mapping. A 4° -sized effective stimulus was presented at 35 positions ranging from 3° ipsilateral to 9° contralateral and from 9° below to 9° above the horizontal meridian. Adjacent positions differed by 3° , horizontally or vertically. The positions were tested interleaved. The number of unaborted presentations per position was at least 4 (mean = 6.5).

Position tolerance test. Five 4° -sized stimuli from different classes (at least two bodies) and ranging from effective to ineffective were presented interleaved at the fovea and at another position (mean eccentricity = 4.6° , minimum distance = 3°) inside the RF. The five images could be from different stimulus classes. The mean number of unaborted presentations per stimulus and location was 11.3 (minimum = 10).

Size tolerance test. An effective and an ineffective image, with sizes of 2° , 4° , and 8° , were presented randomly

interleaved. The location of the stimuli depended on the RF of the neuron: either foveally or the optimal peripheral location. Across neurons, the two images could be either from the same (e.g., both monkey bodies) or different classes. Because no difference in selectivity or tolerance was observed between whether or not the two images were from the same class, we combined the data of all the size tolerance tests in the present paper. The mean number of unaborted presentations per size and image was 11.9 (minimum = 6).

Rotation test. Eight orientations (45° step) of a 4° -sized effective stimulus were shown randomly interleaved at the same location as for the size test. A subset of neurons was also tested with rotated images of the same size as that used in the search test. These tests produced similar results to those obtained with the 4° -sized stimuli. The mean number of unaborted presentations per orientation was 11.8 (minimum = 10).

Silhouette–outline test. We selected 10 images, one of each class, ranging from effective to ineffective, and these were presented randomly interleaved in three rendering conditions: original (same images as in search test), silhouette, and outline versions. The stimulus location was the same as in the size test. For a small proportion of the neurons, the test did not include the outline versions. The mean number of unaborted presentations per stimulus version was 10.1 (minimum = 5).

Data Analysis

Firing rate was computed for each unaborted stimulus presentation in two analysis windows: a baseline window ranging from 100 to 0 msec before stimulus onset and a response window ranging from 50 to 250 msec after stimulus onset. The responses in each test were assessed for statistical significance by a split-plot ANOVA with repeated-measure factor Baseline versus response window and between-trial factor Stimulus condition. A test for a neuron was included only when either the main effect of the repeated factor or the interaction of the two factors was significant ($p < .05$) for that test in that neuron. All further analyses were based on baseline subtracted, average net responses (except for the computation of the separability index; see below). In most statistical analyses (unless otherwise stated) and figures, we employed normalized responses by dividing the response by the maximum of the responses across the different conditions for each neuron. Analyses on net firing rates produced highly similar results.

The degree of body category selectivity was quantified by the body selectivity index (BSI; Popivanov et al., 2014):

$$\text{BSI} = \frac{\bar{R}_{\text{body}} - \bar{R}_{\text{nonbody}}}{|\bar{R}_{\text{body}}| + |\bar{R}_{\text{nonbody}}|}$$

where $\bar{R}_{\text{body}}$ and $\bar{R}_{\text{nonbody}}$ are the mean net firing rates to bodies (monkey bodies, human bodies, mammals, and birds) and nonbodies (faces, objectsM, objectsH, fruits/vegetables) of the search stimulus set, respectively.

The RF size was defined as the square root of the area within the 50% contour line, computed by linear interpolation of the responses at neighboring locations using the Matlab *contourf* function. In many neurons, the 50% contour line was not closed. For those neurons, the boundary of the mapped region was used to close the contour, yielding a minimal, underestimated measure of the RF size. The peak of the RF was defined as the stimulus location with the greatest response.

The main analysis of the tolerance and silhouette–outline tests consisted of a ranking analysis (Sary, Vogels, & Orban, 1993). For each neuron, the stimuli were ranked by their response strength in a reference condition (e.g., foveal position, particular stimulus size, or original version of the stimuli). Then, for each condition, the normalized responses were averaged across neurons as a function of the thus defined stimulus rank (rank 1 = “best” stimulus, etc.). Nonparametric statistical tests (Friedman ANOVA or Wilcoxon matched-pairs test) were employed to assess the significance of the effect of stimulus rank for the non-reference conditions. Only neurons for which the stimulus selectivity for the reference condition was significant (Kruskal–Wallis ANOVA; $p < .05$) were included. Conditions employed as reference are described in Results.

To compute the population orientation tuning curve, we first determined the preferred rotation of each neuron by using the mean response based on half of the trials per rotation condition. Then we plotted the mean response of the other half of the trials using the thus defined preferred orientation. Averaging the normalized responses across neurons was performed after aligning the preferred orientations of the neurons.

To compare the responses to the same object image across size or orientation, we computed for each neuron a Best–Worst index (BWI) defined as the response to the Best size or orientation minus the response to the Worst size or orientation, divided by the response to the Best. The Best and Worst corresponded to the stimuli that evoked the maximum and minimum response, respectively. For the size tolerance test, the index was computed for the most effective of the two images based on the response averaged across the three sizes. To compare for each neuron the responses to the silhouette or outlines and that to the original stimuli, we took the maximal response ($\max(R)$) of the 10 stimuli for each version and then computed the following Responsivity Index (RI),

$$RI = \frac{\max(R_{\text{original}}) - \max(R_{\text{version}})}{|\max(R_{\text{original}})| + |\max(R_{\text{version}})|},$$
 with “version” being the silhouette or outline. To quantify the tolerance for the silhouette and outline transformations, we computed for each neuron the Pearson correlation coefficients r between the responses for the different versions. In addition to these “raw” correlation coefficients, we also computed for each

neuron “noise-corrected” correlation coefficients that were normalized by the reliability of the responses to the different stimulus versions. The Spearman–Brown corrected split-half correlation coefficient for each version was used as measure of the reliability of the neural responses for that stimulus version. The “noise-corrected” correlation coefficients were then computed as the “raw” correlation coefficient divided by the square root of the product of the Spearman–Brown corrected split-half correlation coefficients.

To quantify the tolerance for position or size of a single neuron, we computed a separability index, which is a useful metric of the degree of tolerance of the stimulus preference to image transformations (e.g., Li, Cox, Zoccolan, & Dicarlo, 2009; Brincat & Connor, 2004). First, the gross responses of each cell were tabulated in an $m \times n$ matrix M , with m and n corresponding to the different stimuli (e.g., the five stimuli of the position test) and the transformation variable (e.g., position), respectively. The predicted response, assuming separability of the stimuli and the transformation variable, was then computed as the product of the first principal components of the singular value decomposition of M (see Mysore, Vogels, Raiguel, Todd, & Orban, 2010). The separability index equals the squared Pearson correlation (r^2) between the actual and the predicted responses and could range between 0 (*no separability*) and 1 (*perfect separability*).

We investigated the neural representation of original and silhouette versions of the stimuli with hierarchical cluster analysis of the normalized responses to the 10 original and 10 silhouette stimulus classes. First, we assigned the 10 original and 10 silhouette stimulus versions, which differed across neurons, to the 10 classes. Note that different neurons could be tested with different stimuli of a particular class, but each neuron was tested with one stimulus (in two versions) of each of the 10 classes. In this analysis, we ignored which particular stimulus of a class was presented to the neuron and kept only its class label. Thus, the analysis clusters responses of the different classes and versions (silhouette and original versions), ignoring differences among the exemplars (e.g., different monkey bodies) that belong to the same class. This is sufficient to assess whether responses to a stimulus (e.g., a monkey body) are similar for their original and silhouette versions (e.g., silhouettes and original versions of a monkey body cluster together). After assigning the class labels to each stimulus, we computed the Euclidean distance for all pairwise combinations of the 20 class $\times$ stimulus version combinations, using the normalized net responses of the 117 tested neurons (Op de Beeck, Wagemans, & Vogels, 2001). Finally, we performed a hierarchical cluster analysis (Ward’s method) of the 20×20 distance matrix.

The relationship between the responses in the silhouette–outline test and image properties (e.g., luminance; spectral power) was examined for each neuron by fitting a second order polynomial: $y = ax^2 + bx + c$, with y being the net response strength of the neuron and x the property value

of the 30 images. The second instead of a first order polynomial (linear model) was chosen because it can also capture a tuning to intermediate values of a parameter (Kayaert, Biederman, & Vogels, 2003). The goodness of fit was defined by the square of the Pearson correlation coefficient between the real and the predicted data. A fit was considered to be statistically significant when either the first- or second-order coefficient was outside its 99% confidence interval (using the Matlab function *fit*).

Psychophysical Study

Participants

Nineteen naive human observers (10 women; median age = 28 years, range = 23–57 years) participated. The participants signed an informed consent form.

Stimuli, Apparatus, and Design

The stimuli consisted of 100 silhouette versions of all the objects of the Search test. The stimuli were presented once for 200 msec in random order on a CRT display. The same stimulus presentation, including the noise backgrounds, was employed as in the single-unit study. The participants were instructed to decide whether or not the image depicted a body (excluding faces) by pressing either one of two keys (forced-choice procedure). No feedback was provided. After pressing a key, there was a 1500-msec interval with the noise background still on the display before the next stimulus was presented. They were informed that some bodies had their head cut off and that the proportion of bodies could differ from 50%. Before the test, the participants were given a short 20-presentation practice session in which two original, nonsilhouette versions of stimuli from each of the 10 classes were shown. These stimuli differed from the one employed in the subsequent silhouette test.

RESULTS

Neurons in the mid-STS body patch respond on average more strongly for images containing bodies (monkey bodies, human bodies, mammals, and birds) than to images of other objects, including faces. Here we examine the tolerance of shape-preserving stimulus transformations on the response of these neurons.

Receptive Fields

First we measured the response to an effective stimulus at 35 positions (see Methods) of 175 mid-STS body patch neurons in two monkeys. Figure 1A and B shows the responses of two example neurons in this test. The RF of both neurons included the fovea, which was typical for the population of neurons, and the response depended

on the position of the stimulus inside the RF. The RF of the population of neurons (normalized responses of neurons averaged per position) are shown for monkeys E ($n = 129$ neurons) and B ($n = 46$) separately in Figure 1C and D. Only neurons for which the maximum net firing rate in this test was at least 10 spikes/sec were included. The median maximal net firing rate of the neurons in this test was 42 spikes/sec (first quartile: 29 spikes/sec; third quartile: 70 spikes/sec). In both monkeys, the peak of this population RF corresponded to the fovea. The strong preference for the foveal region is also present when examining the distribution of the peaks of the individual neurons (Figure 1E): 71% of the neurons had their peak within a 6° square centered on the fovea. In addition, for both the average response (Figure 1C, D) and the RF peaks (Figure 1E), there was a marked bias toward the contralateral lower visual field. Note that average net responses in the ipsilateral visual field were positive (Figure 1C, D), with uninterpolated normalized values above 0.1 in each monkey, indicating the presence of bilateral RFs.

Figure 1F shows the distribution of the estimated RF size (square root of the RF area at half maximum; see Methods). The median estimated RF size was 8.3° , with no significant difference in size between the two monkeys (Mann–Whitney U test; $p > .05$). However, because of the (relatively small) extent of the stimulated visual space, the RF size was underestimated in 79% of the neurons (see Methods). Notable is the large variation in RF sizes: from as little as 2° till 14.8° , which is the maximum possible value given our stimulation grid. There was a significant positive correlation between RF size and response strength, defined as the maximum net response across the different positions (Spearman rank correlation, $r_s = .38$; $p < .0001$). However, the range in RF size cannot be merely explained by firing rate, because some neurons with RFs smaller than 5° , which were completely inside the stimulated visual space, had net firing rates well above 50 spikes/sec.

We had a sufficient number of trials per stimulus in the search test (at least five trials) in 114 of the neurons for which we plotted the RF. About half (45%) of these neurons had a BSI greater than 0.33 (twofold stronger responses to bodies compared with nonbodies) and thus were defined as body category selective. The median RF size did not differ significantly between the body category selective (median RF size = 7.4°) and the other neurons (8.6° ; Mann–Whitney U test; $p > .05$).

Position Tolerance

To assess the position tolerance of the stimulus preference, we measured the response of 25 neurons (monkey E: 9 neurons) to five images that were presented at two positions inside the RF. For each neuron, the stimuli were ranked according to the response at one position (the fovea; reference [R] position in Figure 2A), and this stimulus ranking was then used to plot the normalized responses for the other position (NR in Figure 2A). The

Figure 1. RFs of the body patch neurons. (A) Interpolated RF map of an example neuron having a relatively small RF. The nine contour isolines divide the net response in 10 equally spaced color-coded steps. The response at each tested location (small open circles) was measured for the 4° version of the effective stimulus (shown at the right bottom corner). The fovea corresponds to the (0,0) coordinate. VF = visual field. (B) Interpolated RF map for an example neuron with a large RF; same conventions as (A). (C and D) Interpolated average RF maps across all tested neurons for monkey E and monkey B, respectively. Before averaging, the responses for each neuron were normalized by division with the highest net response. (E) Distribution of RF centers across the 35 tested positions. Data were pooled across the two monkeys. (F) RF size distribution, measured as the square root of the area within the contour corresponding to 50% of the maximum net response. Gray bars (estimated) correspond to neurons for which the contour was closed within the mapped area. White bars denote cases in which the contour was open, providing an underestimation of the real RF size. The data of both monkeys were pooled.

neurons showed a marked position tolerance of their stimulus preference: The mean normalized response at position NR, which was not used for the ranking, declined monotonically as a function of stimulus rank, which was

significant in each monkey (Friedman ANOVA of ranked responses at position NR; $p < 10^{-5}$). As a metric of the degree of position tolerance, we computed the separability index for each neuron (see Methods). The median

Figure 2. Position and size tolerance. (A) Average ranked normalized net responses to five effective and ineffective stimuli, measured in the position tolerance test. The responses at the fovea, plotted in gray stippled line, were used for ranking (R), and the responses at the other position (NR) were plotted using the same rank. Note that the neurons we examined showed sizable responses to the “best” stimulus at the ranked position, the average net firing rate being 38 spikes/sec. (B) Distribution of the separability indices of the neurons recorded in the position tolerance test. (C) Average ranked normalized net responses to the effective and ineffective stimuli at three different sizes. The largest size (8°) was used to determine the stimulus rank (light gray stippled line). The response at the other two sizes, 4° (dark gray solid line) and 2° (black solid line), were plotted using the same rank. The average net firing rate to the best stimulus for the ranked size was 40 spikes/sec. (D) Distribution of the separability indices of the neurons recorded in the size tolerance test. The data in all panels are pooled across the two monkeys. Error bars denote the *SEM*.

separability index was 0.95 (quartiles: 0.88 and 0.98, $n = 25$; Figure 2B). These high separability indices, in combination with the ranking analysis of Figure 2A, indicate that the stimulus preference of the large majority of neurons was well preserved across the tested positions.

Size Tolerance

The size tolerance was assessed in 156 neurons (99 neurons in monkey E) by presenting two images at three different sizes (range = 2 octaves). In most neurons, the

response varied little with size: The median BWI (see Methods) was 0.43 (first quartile: 0.28; third quartile: 0.65), indicating less than a twofold variation (including response variability) in response with size. The size eliciting the highest response was 8° in 47% of the neurons. The two smaller sizes were preferred equally (25% and 28%). The size associated with the smallest response was 2° in 53% of the neurons. There was no significant correlation between the BSI and the BWI (Spearman rank correlation, $r_s = .15$; $p > .05$; $n = 135$).

To assess the size tolerance of the stimulus preference, we ranked the images according to their response strength for the 8° size and then plotted the normalized response for the thus ranked images for the other two sizes. This analysis was performed for those neurons ($n = 137$) that showed a significant selectivity for the 8° stimuli. As shown in Figure 2C, even changing the size by two octaves preserved the shape preference, which was significant in each animal (Wilcoxon matched-pairs tests; $p < 10^{-8}$). Similar results were obtained when the reference condition for the ranking was 4°. In agreement with these ranking analyses, the median separability index was very high ($r^2 = .99$, quartiles: .96 and 1; Figure 2D), showing that the stimulus preference was well preserved across size.

In-plane Orientation Tuning

Popivanov et al. (2014) showed a strong effect of rotation of isolated body parts on the responses of mid-STS body patch neurons. Here, we examined the effect of rotation for 4°-sized whole bodies with (mammals and birds) or without head (monkey and human bodies), preserving the structural relationship among the individual body parts. Figure 3A shows the responses to an effective stimulus as a function of rotation for an example neuron, demonstrating a strong tuning for in-plane orientation. This strong tuning for rotation was typical for the population of tested neurons ($n = 40$ neurons; monkey E: 19), as is evident from the population orientation tuning curve (Figure 3B). In fact, a 45° rotation from the preferred orientation ("best") was sufficient to reduce the response by almost 50% in each animal. For each neuron, we computed a BWI (see Methods) comparing the response to the best and the worst rotation. The distribution of the BWI (Figure 3C) was strongly skewed toward high values, with a median of 1.03. In fact, for 75% of the neurons the BWI was ≥ 1 , indicating no response or inhibition to at least one rotation. In 36 of the 40 tested neurons, the effective stimulus was a body, whereas the other four neurons were tested with a face (2), a sculpture, or a manmade object. Because the orientation tuning was equally strong for the body (median BWI = 1.03) compared with the nonbody stimuli (median BWI = 1.17), we pooled the neurons in the above analysis.

In 18 neurons (monkey E: 5), we tested the effect of rotation both for the 4° image and the original scaled

Figure 3. In-plane orientation selectivity. (A) A polar plot showing the orientation tuning of an example neuron. The 0° rotation condition corresponds to the image rotation employed in the search test and also is the "best" rotation condition for this neuron. The images (size 4°) are shown next to the plot. The BWI of the neuron is indicated above the plot. (B) Population orientation tuning curve, shown separately for the neurons of monkey E (blue) and monkey B (red). Before averaging, the tuning curves were aligned according to the best orientation per neuron (0° = "best"), based on the responses in half of the trials. The averaging across neurons was done for the other half of the trials, using net responses normalized per neuron. The average firing rate for the best orientation was 60 and 41 spikes/sec for monkeys E and B, respectively. (C) The distribution of BWI for all tested neurons, pooled across monkeys. White bars correspond to body category-selective neurons (BSI ≥ 0.33), whereas the gray bars denote cells having a BSI < 0.33 in the main test. The arrows indicate the medians of the two groups of cells.

image that was employed in the search test. For both stimulus scales, we observed an equally strong effect of rotation. Moreover, the responses to the eight rotations correlated strongly between the two scales (median Pearson $r = .91$; $\min(r) = .47$; $n = 18$), indicating that the effect of rotation does not interact with image size. The latter lessens the concern that the rotation effects result from an RF inhomogeneity.

Stimulus Reduction: Selectivity for Silhouettes and Outlines

The images that were presented in all the tests thus far were achromatic and included both shading and textures,

that is, information about material properties. Material properties, such as presence or absence of fur, skin, feathers, and so forth, may contribute to responses of the body patch neurons. To assess the contribution of material properties and shading, we reduced the original stimulus to a black silhouette and an outline version. Note that these stimulus transformations preserve the two-dimensional shape of the image. Figure 4 shows the responses of two example neurons to each of the three versions of 10 images that belong to different classes. The neuron shown in Figure 4A shows similar responses and selectivity to the original and silhouette versions. The responses to the most effective stimulus was strongly reduced for the outline compared with the original and

Figure 4. Responses to originals, silhouettes, and outlines compared: single neuron examples. Ranked responses to original (light gray stippled line), silhouette (black solid line), and outline (dark gray solid line) versions of the stimuli (images shown at the bottom of each panel). The stimuli of all versions were ranked according to the response to the original version (from best to worst). The two example neurons show a high positive (A) or a low negative (B) correlation between the responses for the three versions of the stimuli (original, silhouette, and outline). $r(\text{Originals}; \text{Silhouettes})$ or $r(\text{Originals}; \text{Outlines})$ correspond to the Pearson correlation coefficients for the responses to Originals and Silhouettes, or Originals and Outlines, respectively. Error bars denote the *SEM*.

silhouette, but the overall preference was largely preserved. The neuron of Figure 4B shows a highly different response profile for the three stimulus versions: highly different preferences for the original compared with the silhouette versions and little if any responses to the outlines. Such neurons were a minority in the population of neurons that we recorded.

We compared the responses to the original and silhouette versions in 117 neurons (monkey E: 64 neurons), 114 of which showed a significant selectivity for the 10 original stimuli. For each of the latter neurons, we ranked the 10 images according to their response, and this rank was employed to plot the responses to the silhouettes (Figure 5A, left). The average normalized response to the silhouette decreased with stimulus rank (full line in Figure 5A, left), and this was significant in each animal (Friedman ANOVA; both p s < .0001). Thus, there was a strong tolerance for this shape-preserving transformation.

To compare the best responses to the original and silhouette images, we computed the RI (see Methods) for the 113 neurons that showed a significant response to either one of the two versions and also for which we had sufficient trials in the search test to compute a BSI. The latter allowed us to determine whether the effect of the silhouette transformation correlates with body category selectivity of the neuron. A positive RI corresponds to a stronger response to the original compared with the silhouette. The median RI of all neurons was -0.04 (not significantly different from 0, $p > .05$; Figure 5B left), indicating that overall the response to silhouettes and originals were rather similar. This is remarkable because we did not attempt to equate the luminance of the two stimulus versions. About half of the neurons responded more to the silhouette than to the original image, and only 11% of the neurons responded at least twice as strong to the original than to the silhouette ($RI > 0.33$), indicating that shape features alone can drive rather well the large majority of mid-STS body patch neurons. There was no correlation between the BSI and the RI (Spearman rank correlation, $r_s = .02$; $p > .05$; Figure 5B, left).

The ranking analysis presented in Figure 5A shows the tolerance averaged across neurons. To obtain a tolerance estimate for each individual neuron, we computed the Pearson correlation coefficient between the responses to the 10 images for the original and silhouette versions. This correlation was .96 and $-.07$ for the single neuron of Figure 4A and B, respectively. The median correlation for the whole analyzed population of neurons was .63 (Figure 5C, left), which is significantly higher than 0 (Wilcoxon signed-rank test $p < 10^{-18}$; $n = 113$), and there was no significant difference between the two monkeys (.65 vs. .60). The magnitude of these correlations depends on both the relationship between the responses to the two versions and the reliability of these responses for each of the versions. To correct for the contribution of the latter noise factor to the correlations, we also computed for each neuron “noise-corrected” correlation

coefficients (see Methods). The median “noise-corrected” correlation coefficient was .75 (first quartile = .50; third quartile = .92), which implies that the responses to shape features account for 56% of the explainable variance of the responses to the original images.

Interestingly, there was a weak but significant correlation between the BSI and the correlation between the responses to the two stimulus versions (Spearman rank correlation, $r_s = .37$; $p < .0005$), and this was true in each animal ($r_s = .35$ and .37). Indeed, body category-selective neurons showed on average higher correlations between their responses to the two stimulus versions (median $r = .79$; $n = 55$) than the other neurons (median $r = .50$; $n = 58$; Mann–Whitney U test; $p < .0004$). This difference between the groups of neurons was also present for the “noise-corrected” correlation coefficients (Mann–Whitney U test; $p < .05$), with median values of .87 and .65 for the body category-selective and nonselective neurons, respectively. Thus, shape accounted for 77% of the explainable variance for the “average” body category-selective neuron.

The responses to the outline versions were on average lower than the original (median RI = 0.25; Wilcoxon signed-rank test; $p < 10^{-9}$; $n = 84$; Figure 5B, right), with 38% of the neurons having an RI of larger than 0.33. Despite the lower response, both the ranking (Figure 5A, right; Friedman ANOVA: $p < .0001$) and the correlation analyses (Figure 5C, right; median correlation = .43; Wilcoxon signed-rank test, $p < 10^{-12}$) showed some preservation of stimulus preference when comparing the original and outlines. However, the degree of tolerance to the outline transformation (median “noise-corrected” correlation coefficient = .54, corresponding to 29% of explainable variance) was markedly lower than for the silhouette transformation. This was especially true for the body category-selective neurons, with both groups of neurons showing a similarly low tolerance for the outline transformations (Figure 5C, right).

The high tolerance to the silhouettes of the mid-STS body patch neurons was corroborated at the population level by performing cluster analysis of the normalized responses to the 10 original and 10 silhouette stimulus classes (see Methods). Indeed, the hierarchical cluster analysis showed a tight clustering of the original and silhouette versions for each class (Figure 6A). Also, the analysis shows a clustering of monkey bodies and four-legged mammals, of human bodies and birds, and a third cluster that consisted of the nonbody classes, including faces. However, caution is needed when interpreting the clustering of the different classes, because, first, only one stimulus per class was tested in each neuron and, second, online selection of the stimuli based on their response may have introduced biases. However, note that the clustering of the original and silhouette versions within a class cannot be explained by such potential selection biases, because these versions were defined a priori.

Figure 5. Responses to originals, silhouettes, and outlines compared: population analysis. (A) Mean ranked normalized responses to the original and silhouette versions (left) and to the original and outline versions (right). Before averaging the net responses across neurons, stimuli were ranked by the response to the original versions. Error bars denote the *SEM* and *n* correspond to the number of analyzed neurons. Data of both monkeys are pooled. The average net firing rate for the original “best” stimulus was 41 spikes/sec. (B) Distributions of the RI for original-silhouette comparisons (left) and original-outline comparisons (right). White and gray bars correspond to the body selective or the other cells, respectively (same conventions as Figure 3C). (C) The left panel plots the Pearson correlation between the responses to original and silhouette versions as a function of the BSI. The right panel shows the same for the correlation between the original and outline versions. Blue and red crosses correspond to the single neurons of monkeys E and B, respectively. The two highlighted points (*a* and *b*) denote the example neurons shown in Figure 4, and the letters correspond to the panels of Figure 4. Distributions of the Pearson correlation coefficients are shown to the right of the scatter plot, with the white bars indicating the body category-selective neurons (same conventions as B). Arrows indicate medians.

Figure 6. Responses to silhouettes for body patch neurons and by human participants. (A) Hierarchical cluster analysis tree of the normalized net responses to original and silhouette versions of the stimuli, grouped per class (human bodies, birds, fruits/vegetables, objectsH, sculptures, objectsM, human faces, monkey faces, mammals, and monkey bodies) and version (original images and their silhouettes). Each combination of class and stimulus version is indicated by an example image at the bottom of the panel. Note that the responses to the silhouette and original are clustered per class. (B) Matrix showing the behavioral response “body” (black) to the silhouette versions of each of the 100 stimuli for each of the 19 human participants. Stimuli are ordered per class, as shown by the representative images at the bottom of the panel. The red vertical line separates body from nonbody stimuli. The five most confused stimuli are shown on the top of the panel with arrows pointing to their columns in the matrix.

To examine the contribution of low-level image properties to the responses for the 30 images of the silhouette–outline tests, we fitted second-order polynomials to the responses as a function of a set of low-level image properties (see Methods): mean luminance, contrast, full power spectrum, power spectrum along the horizontal or vertical axis, and the power spectrum along each of the two orientation axes for low and high spatial frequencies separately. For each of the nine regressions per neuron, the fits were poor, with the maximum median r^2 being .08. Few neurons showed statistically significant fits, the highest number being 9 (11%), which was obtained when regressing the power in either one of the two orientation bands. This analysis suggests that, overall, low-level properties such as luminance, contrast, and spectral power contributed little to the variation in responses across stimuli.

Categorization of Silhouettes as Bodies or Nonbodies: Human Psychophysics

Given the high degree of tolerance of the mid-STS body patch neurons, especially of the body category-selective

ones, to the shape-preserving silhouette transformation and the assumption that these neurons contribute to body versus nonbody categorization (Popivanov et al., 2014), we wondered whether the shape information present in silhouettes indeed is sufficient to categorize bodies. To answer this question, we assessed how well naive human participants were able to categorize the silhouette stimuli, which we employed in the macaque single-unit study, as bodies or nonbodies (see Methods). The mean categorization performance for the silhouettes was 91% correct ($n = 19$ participants) with 89% and 7% of “body” decisions for the body and nonbody silhouettes, respectively. Further analysis of the behavioral data showed that the categorization errors were not entirely random (Figure 6B). Three of 40 body silhouettes were wrongly categorized by more than 7 of the 19 participants (Binomial test; $p < .01$ with expected $p = .11$): two monkey bodies (13/19 and 10/19 errors) and a frontal view of a beaver (15/19). Two of 60 nonbody silhouettes were wrongly categorized as a body by more than 6/19 participants (Binomial test; $p < .01$ with expected $p = .07$): a human face (8/19) and a corn (11/19).

Popivanov et al. (2014) showed that the population of body patch neurons—which overlapped with the sample studied here—distinguished bodies from nonbody images (including faces). These stimuli were the original versions from which the silhouettes of this study were derived. The clustering of the neuronal data was imperfect, and thus, we assessed whether the categorization errors, based on the clustering of the monkey single-unit responses for the original images, match those for the silhouettes in the human psychophysics. Only one body was misclassified by the population of neurons, and this was a human body, which does not fit the human behavioral data for the silhouettes. The body cluster of the neurons contained five nonbodies and two of these corresponded to the two nonbody silhouettes wrongly classified by the human participants. It is very unlikely that the latter occurred by chance ($p = .0054$).

DISCUSSION

Mid-STs body patch neurons respond on average more strongly to images of bodies compared with nonbodies but also show a strong selectivity for different body images (Popivanov et al., 2014). Here we demonstrate that the selectivity of mid-STs body patch neurons shows a high degree of tolerance to changes in the position and size of the object, but their response strongly depends on the in-plane orientation of a body. Importantly, the selectivity of the majority of these neurons was to a large degree preserved when silhouettes of the images were presented, suggesting that mainly shape-based features drive these neurons. In a subsequent human psychophysical study, we showed that indeed the information present in silhouettes is largely sufficient for body versus nonbody categorization. Remarkably, the few nonbody silhouettes that were classified as bodies by the human observers were also misclassified by a population of mid-STs body neurons (Popivanov et al., 2014), in agreement with the dominant role of shape features in driving the response of these neurons.

The RFs of the mid-STs body patch neurons differed in three respects from those reported for anterior IT by Op de Beeck and Vogels (2000): The population RF of the body patch neurons was smaller, their average responses to ipsilateral stimuli were less, and they showed a bias toward the lower visual field. The smaller RF and weaker ipsilateral responses fit the more posterior location of the mid-STs body patch. Our results appear to be in line with Hikosaka (1998), who showed low eccentricity RFs in the ventral bank of this part of the STs in the anesthetized *Macaca fuscata*, with a lower visual field bias in one of his two animals. We would like to note that the bias toward RFs with small eccentricities in our sample of neurons is at least partially because of the fact that the body patch was defined with a fMRI study with foveally centered stimuli (Popivanov et al., 2012). Thus,

it is well possible that body-selective neurons are also present in regions with more eccentric RFs.

Interestingly, also the extrastriate body area (EBA), defined in human fMRI by contrasting headless bodies versus other objects, is more strongly activated for stimuli located in the lower compared with the upper visual field (Schwarzlose, Swisher, Dang, & Kanwisher, 2008). Why would a body patch show a lower visual field bias? As also pointed out by Schwarzlose et al. (2008), this bias may reflect the tendency of primates to foveate more frequently faces when looking at bodies (Yun, Peng, Samaras, Zelinsky, & Berg, 2013; Shepherd, Steckenfinger, Hasson, & Ghazanfar, 2010), inducing a lower field bias in experiencing body parts below the head. This visual field bias in where body parts are experienced could then produce the position-dependent bias in the responses of the neurons. This is not an unlikely possibility, because fMRI activations in the EBA reflect the experienced location of body configurations (Chan, Kravitz, Truong, Arizpe, & Baker, 2010). Another potential explanation is based on the sort of stimulus features that may drive the mid-STs body patch neurons. The responses to stimuli like the face shown in Figure 6B suggest that limb-like shape features drive body patch neurons well (see Popivanov et al., 2014), although this conjecture needs formal testing. This together with the fact that the majority of the monkey body images that we employed to map the body patch contained limbs in the bottom part of the image may also explain the lower visual field bias. Also note that four-legged mammals and birds, stimuli that activate the body patch well (Popivanov et al., 2012), have such extended shape features predominantly in the lower part of their body.

How do the RFs of the mid-STs body patch compare to those of the neighboring face patches? The more posterior face patch PL appears to have smaller RFs with a median of approximately 4° (Figure 7A in Issa & Dicarlo, 2012), which is about half of the (underestimated) median size in the body patch. Another difference with the mid-STs body patch is that the RFs in PL are biased towards the upper instead of the lower visual field (Issa & Dicarlo, 2012). This upper field bias seems to be related to the location of an eye in the faces employed to map PL (Issa & Dicarlo, 2012). Hence, it is possible that mapping face-selective regions with faces presented at other retinotopic positions may produce other, neighboring face-selective PL-like patches, corresponding to the rough retinotopy of this posterior part of IT (Boussaoud, Desimone, & Ungerleider, 1991). Freiwald and Tsao (2010) measured the responses of face-selective neurons in ML, AL, and AM to stimuli at four eccentricities, ranging between 0° and 13° . On the basis of the averaged responses across neurons shown in their Figure S10, RF size increased with the anterior level of the face patch. The average response in ML, which is located close to the mid-STs body patch, at 6° eccentricity was about half of the response at the fovea, suggesting RFs comparable in size to those of the

mid-STS body patch. However, in the Freiwald & Tsao, 2010 study, images were large (8°) and responses were averaged across ipsi- and contralateral visual fields, precluding a definite comparison between studies. Reverse correlation mapping of ML multiunits showed, as for PL, a bias toward the contralateral eye region of a 6° face (Issa & Dicarlo, 2012), suggesting at least a strong contralateral bias in ML.

An increasingly growing literature shows that the stimulus preference of macaque IT neurons tolerates at least moderate changes in stimulus location and size (Dicarlo et al., 2012; for a review of older studies, see Vogels & Orban, 1996). We show here that the same holds for the mid-STS body patch, implying that the output of these neurons can be used to discriminate bodies irrespectively of the size and position of the stimulus (Li et al., 2009). Note, however, that, similar to other IT neurons (Vogels & Orban, 1996), the stimulus preferences were tolerant to size and position but not to the response strength.

Extending our previous observation that mid-STS body patch neurons were sensitive to the orientation of single body parts (Popivanov et al., 2014), we now show that these neurons are highly selective for the in-plane orientation of whole body images. This implies that the orientation selectivity for body parts was not because of the absence of a reference frame, that is, other parts of the body. Even with configuration information present, the orientation selectivity was pervasive. A study in which recordings without fMRI guidance were performed from STS neurons selective for human bodies also reported orientation-selective responses to human bodies in 22 of 26 neurons (Ashbridge, Perrett, Oram, & Jellema, 2000). However, the orientation selectivity in that study was less than what we observed in the mid-STS body patch, which may reflect the more anterior location of the neurons in that study compared with ours or different functional properties of neurons inside versus outside the body patches. Ashbridge et al. (2000) noted that most neurons preferred the upright orientation, which was also the case in our study (data not shown). This apparent preference for upright bodies may be related to the “body inversion effect” in human perception (Reed, Stone, Bozova, & Tanaka, 2003) and event-related potentials (Minnebusch, Keune, Suchan, & Daum, 2010) but also may merely reflect the fact that in both studies the neurons were searched using upright bodies, that is, a simple search bias.

The responses of the majority of mid-STS body patch neurons to static images of objects tolerated the deletion of texture and shading features, implying that shape is a critical feature that drives these neurons. The comparable maximal responses to the original and silhouettes found here for the macaque mid-STS body patch fits fMRI studies of the EBA that showed similar activations for human bodies and their silhouettes (Peelen & Downing, 2007; Downing et al., 2001).

The preserved selectivity for body silhouettes in the mid-STS body patch contrasts with the importance of

internal face features, including contrast polarity, for the face selectivity in the neighboring face patch ML (Yue, Nasr, Devaney, Holt, & Tootell, 2013; Ohayon et al., 2012; Freiwald et al., 2009). This points to a fundamental difference between the stimulus processing in the body and face patches despite their proximity in location in the cortex. Face detection as well as face discrimination rely strongly on internal features (Dakin & Watt, 2009), whereas as our psychophysical study shows, body versus non-body categorization is relatively immune to the deletion of internal features. This difference between face perception and body categorization dovetails with the properties of the neurons in the mid-STS body and face patches. It still remains to be studied how these two systems processing stimuli in a different way interact to obtain a holistic body + face percept. There is a strong possibility that neurons that are located between the two patches and which respond to both bodies and faces (Popivanov et al., 2014) play an important role in integrating the information from the two patches, which corresponds with the smooth gradients in category selectivity in the human brain reported recently (Huth, Nishimoto, Vu, & Gallant, 2012).

Does this mean that internal features, material properties, or shading play no role in body processing? Our data show that the correspondence between the preference for the original and silhouettes was not perfect for some body-selective neurons, even when taking into account the reliability of the responses. Thus, some of these neurons may show some sensitivity to material properties (Koteles, De Maziere, Van Hulle, Orban, & Vogels, 2008). Another not mutually exclusive possibility is that the shading or textural cues assist in the segmentation of partially overlapping body parts (Vogels & Biederman, 2002; Missal, Vogels, & Orban, 1997). However, the high degree of tolerance of the single units to the silhouette transformation suggests that, overall, these textural and shading cues play only a secondary role in the feature selectivity in this body patch.

As in face selectivity studies (Tsao et al., 2006) and for easier control of low-level stimulus parameters, we employed achromatic stimuli. This choice also was motivated by behavioral work in humans and monkeys that showed only a negligible effect of removal of the color information from scenes on fast animal versus nonanimal categorization (Delorme, Richard, & Fabre-Thorpe, 2000). One could speculate that the importance of color and material properties for body representations is strongly task dependent: These cues likely contribute little to body detection and discrimination, for which shape features are dominant but may inform attractiveness judgments (Cook & Duchaine, 2011) and sexual preferences (Waite, Gerald, Little, & Kraisselburd, 2006). Also, the role of shading and texture cues is likely to be task dependent. For instance, these cues can contribute to the discrimination of a frontal versus a backward view of a body, which produces highly similar silhouettes but are less important for body versus nonbody categorization. Further work is needed to clarify

the relative contributions of the body patches or other regions (e.g., color patches or regions involved in 3-D surface coding) to the different tasks that can be performed on images of bodies.

The present findings show that mid-STS body patch neurons show a strong tolerance for changes in retinal position, size, and the silhouette transform, all shape-preserving transformations. However, they do show strong selectivity for orientation. Shape is classically defined as an orientation-invariant property of an object, indicating that these neurons do not respond to shape per se but to oriented shape features, as in other parts of IT (Brincat & Connor, 2004; Tanaka, 1996). Thus, overall our data suggest that mid-STS body patch neurons respond to oriented, shape features that are more prevalent in images of bodies than of other objects. This explains why some of these neurons also respond to images of nonbodies (e.g., the face and the corn of Figure 6B) that contain limb-like features in their outer contour (Popivanov et al., 2014). Indeed, silhouettes of these images also tend to be classified as bodies by human participants. Our data also show that the shape feature signals of these neurons show a strong position and size tolerance, which can assist position- and retinal size-invariant body categorization and discrimination based on shape. It remains to be seen whether and how neurons in more anterior body patches extend this analysis of body features.

Acknowledgments

This study was supported by the Fonds voor Wetenschappelijk Onderzoek (FWO) Vlaanderen, GOA, IUAP, and PF grants. We are grateful to M. Docx, I. Puttemans, C. Ulens, P. Kayenbergh, G. Meulemans, W. Depuydt, S. Verstraeten, and M. De Paep for technical support; Dr. P. Downing and Dr. M. Tarr for providing some of the stimuli; and Dr. J. Taubert for reading an earlier version of the manuscript. I. D. P. was supported by a fellowship from the Agentschap voor Innovatie door Wetenschap en Technologie (Grant 101071), and J. J. is postdoctoral fellow supported by FWO Vlaanderen.

Reprint requests should be sent to Rufin Vogels, Laboratorium voor Neuro- en Psychofysiologie, O&N2, Campus Gasthuisberg, Herestraat 49, bus 1021, 3000 Leuven, Belgium, or via e-mail: Rufin.vogels@med.kuleuven.be.

REFERENCES

- Ashbridge, E., Perrett, D. I., Oram, M. W., & Jellema, T. (2000). Effect of image orientation and size on object recognition: Responses of single units in the macaque monkey temporal cortex. *Cognitive Neuropsychology*, *17*, 13–34.
- 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. *Journal of Neurophysiology*, *101*, 688–700.
- Boussaoud, D., Desimone, R., & Ungerleider, L. G. (1991). Visual topography of area TEO in the macaque. *Journal of Comparative Neurology*, *306*, 554–575.
- Brincat, S. L., & Connor, C. E. (2004). Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nature Neuroscience*, *7*, 880–886.
- Chan, A. W. Y., Kravitz, D. J., Truong, S., Arizpe, J., & Baker, C. I. (2010). Cortical representations of bodies and faces are strongest in commonly experienced configurations. *Nature Neuroscience*, *13*, 417–418.
- Cook, R., & Duchaine, B. (2011). A look at how we look at others: Orientation inversion and photographic negation disrupt the perception of human bodies. *Visual Cognition*, *19*, 445–468.
- Dakin, S. C., & Watt, R. J. (2009). Biological “bar codes” in human faces. *Journal of Vision*, *9*, 1–10.
- Delorme, A., Richard, G., & Fabre-Thorpe, M. (2000). Ultra-rapid categorisation of natural scenes does not rely on colour cues: A study in monkeys and humans. *Vision Research*, *40*, 2187–2200.
- Dicarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, *73*, 415–434.
- Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science*, *293*, 2470–2473.
- Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*, 598–601.
- Freiwald, W. A., & Tsao, D. Y. (2010). Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science*, *330*, 845–851.
- Freiwald, W. A., Tsao, D. Y., & Livingstone, M. S. (2009). A face feature space in the macaque temporal lobe. *Nature Neuroscience*, *12*, 1187–1196.
- Goris, R. L. T., & Op de Beeck, H. P. (2009). Neural representations that support invariant object recognition. *Frontiers in Computational Neuroscience*, *3*, 1–16.
- Hikosaka, K. (1998). Representation of foveal visual fields in the ventral bank of the superior temporal sulcus in the posterior inferotemporal cortex of the macaque monkey. *Behavioural Brain Research*, *96*, 101–113.
- Huth, A. G., Nishimoto, S., Vu, A. T., & Gallant, J. L. (2012). A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron*, *76*, 1210–1224.
- Issa, E. B., & Dicarlo, J. J. (2012). Precedence of the eye region in neural processing of faces. *Journal of Neuroscience*, *32*, 16666–16682.
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*, 4302–4311.
- Kayaert, G., Biederman, I., & Vogels, R. (2003). Shape tuning in macaque inferior temporal cortex. *Journal of Neuroscience*, *23*, 3016–3027.
- Koteles, K., De Maziere, P. A., Van Hulle, M., Orban, G. A., & Vogels, R. (2008). Coding of images of materials by macaque inferior temporal cortical neurons. *European Journal of Neuroscience*, *27*, 466–482.
- Li, N., Cox, D. D., Zoccolan, D., & Dicarlo, J. J. (2009). What response properties do individual neurons need to underlie position and clutter “invariant” object recognition? *Journal of Neurophysiology*, *102*, 360–376.
- Minnebusch, D. A., Keune, P. M., Suchan, B., & Daum, I. (2010). Gradual inversion affects the processing of human body shapes. *Neuroimage*, *49*, 2746–2755.
- Missal, M., Vogels, R., & Orban, G. A. (1997). Responses of macaque inferior temporal neurons to overlapping shapes. *Cerebral Cortex*, *7*, 758–767.
- Mysore, S. G., Vogels, R., Raiguel, S. E., Todd, J. T., & Orban, G. A. (2010). The selectivity of neurons in the macaque

- fundus of the superior temporal area for three-dimensional structure from motion. *Journal of Neuroscience*, *30*, 15491–15508.
- Ohayon, S., Freiwald, W. A., & Tsao, D. Y. (2012). What makes a cell face selective? The importance of contrast. *Neuron*, *74*, 567–581.
- Op de Beeck, H., & Vogels, R. (2000). Spatial sensitivity of macaque inferior temporal neurons. *Journal of Comparative Neurology*, *426*, 505–518.
- Op de Beeck, H., Wagemans, J., & Vogels, R. (2001). Inferotemporal neurons represent low-dimensional configurations of parameterized shapes. *Nature Neuroscience*, *4*, 1244–1252.
- Peelen, M. V., & Downing, P. E. (2007). The neural basis of visual body perception. *Nature Reviews Neuroscience*, *8*, 636–648.
- Pinsk, M. A., Arcaro, M., Weiner, K. S., Kalkus, J. F., Inati, S. J., Gross, C. G., et al. (2009). Neural representations of faces and body parts in macaque and human cortex: A comparative fMRI study. *Journal of Neurophysiology*, *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. *Proceedings of the National Academy of Sciences, U.S.A.*, *102*, 6996–7001.
- 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. *Journal of Neuroscience*, *34*, 95–111.
- Reed, C. L., Stone, V. E., Bozova, S., & Tanaka, J. (2003). The body-inversion effect. *Psychological Science*, *14*, 302–308.
- Sary, G., Vogels, R., & Orban, G. A. (1993). Cue-invariant shape selectivity of macaque inferior temporal neurons. *Science*, *260*, 995–997.
- Sawamura, H., Orban, G. A., & Vogels, R. (2006). Selectivity of neuronal adaptation does not match response selectivity: A single-cell study of the fMRI adaptation paradigm. *Neuron*, *49*, 307–318.
- Schwarzlose, R. F., Swisher, J. D., Dang, S., & Kanwisher, N. (2008). The distribution of category and location information across object-selective regions in human visual cortex. *Proceedings of the National Academy of Sciences, U.S.A.*, *105*, 4447–4452.
- Shepherd, S. V., Steckenfinger, S. A., Hasson, U., & Ghazanfar, A. A. (2010). Human-monkey gaze correlations reveal convergent and divergent patterns of movie viewing. *Current Biology*, *20*, 649–656.
- Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience*, *19*, 109–139.
- Tsao, D. Y., Freiwald, W. A., Knutsen, T. A., Mandeville, J. B., & Tootell, R. B. (2003). Faces and objects in macaque cerebral cortex. *Nature Neuroscience*, *6*, 989–995.
- Tsao, D. Y., Freiwald, W. A., Tootell, R. B., & Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science*, *311*, 670–674.
- Vogels, R., & Biederman, I. (2002). Effects of illumination intensity and direction on object coding in macaque inferior temporal cortex. *Cerebral Cortex*, *12*, 756–766.
- Vogels, R., & Orban, G. A. (1996). Coding of stimulus invariances by inferior temporal neurons. *Progress in Brain Research*, *112*, 195–211.
- Waite, C., Gerald, M. S., Little, A. C., & Kraiselburd, E. (2006). Selective attention toward female secondary sexual color in male rhesus macaques. *American Journal of Primatology*, *68*, 738–744.
- Yue, X. M., Nasr, S., Devaney, K. J., Holt, D. J., & Tootell, R. B. H. (2013). fMRI analysis of contrast polarity in face-selective cortex in humans and monkeys. *Neuroimage*, *76*, 57–69.
- Yun, K., Peng, Y. F., Samaras, D., Zelinsky, G. J., & Berg, T. L. (2013). Exploring the role of gaze behavior and object detection in scene understanding. *Frontiers in Psychology*, *4*, 1–14.