Body Patches in Inferior Temporal Cortex Encode Categories with Different Temporal Dynamics

Satwant Kumar* and Rufin Vogels

Abstract

■ An unresolved question in cognitive neuroscience is how representations of object categories at different levels (basic and superordinate) develop during the course of the neural response within an area. To address this, we decoded categories of different levels from the spiking responses of populations of neurons recorded in two fMRI-defined body patches in the macaque STS. Recordings of the two patches were made in the same animals with the same stimuli. Support vector machine classifiers were trained at brief response epochs and tested at the same or different epochs, thus assessing whether category representations change during the course of the response. In agreement with hierarchical processing within the body patch network, the posterior body patch mid STS body (MSB) showed an earlier onset of categorization compared with

the anterior body patch anterior STS body (ASB), irrespective of the categorization level. Decoding of the superordinate body versus nonbody categories was less dynamic in MSB than in ASB, with ASB showing a biphasic temporal pattern. Decoding of the ordinate-level category human versus monkey bodies showed similar temporal patterns in both patches. The decoding onset of superordinate categorizations involving bodies was as early as for basic-level categorization, suggesting that previously reported differences between the onset of basic and superordinate categorizations may depend on the area. The qualitative difference between areas in their dynamics of category representation may hinder the interpretation of decoding dynamics based on EEG or MEG, methods that may mix signals of different areas.

INTRODUCTION

An important question in current visual neuroscience is how objects are processed over the course of their presentation: Does the object representation change over the course of the neural response? Recent EEG and MEG studies in humans examined the dynamics of the encoding of stimuli of various visual categories (reviewed by Contini, Wardle, & Carlson, 2017). A common finding in these studies was the poor decoding of a visual category by a classifier when training and test epochs differed in time compared with identical training and test epochs. This suggested that the encoding of visual categories is highly dynamic, with markedly different object representations over the course of the response. However, because of the low spatial resolution of MEG, it is unclear whether this apparent dynamic encoding reflects different category representations of a population of neurons over the temporal course of the response and/or the various contributions of different areas with different stimulus representations to the MEG signal at successive time epochs. Answering this question requires high-spatial-resolution recordings in a visual area, as

offered by single-unit recording. Previous studies using spiking activity recordings in the inferior temporal (IT) cortex of macaque monkeys provided discrepant answers to this question: One study showed dynamic encoding of learned artificial visual categories (Meyers, Freedman, Kreiman, Miller, & Poggio, 2008), whereas others showed a more static encoding of individual objects (Kumar, Kaposvari, & Vogels, 2017; Zhang et al., 2011). Whether this discrepancy was because the recordings were made from different subregions of the IT cortex, different tasks, or the level of categorization is unclear.

Here, we address these gaps in our understanding of the moment-by-moment representation of visual categories in the IT cortex. We systematically examined the dynamics of the encoding of visual categories by employing a decoding approach similar to those of the above human MEG and monkey IT single-unit studies. We compared decoding dynamics between different levels of categorization (superordinate and basic [ordinate]) and between two category-selective patches within IT. The time courses of basic level versus superordinate classification have been controversial, with some studies reporting an early basic level and a later superordinate categorization in macaque IT (Dehaqani et al., 2016) and human MEG (reviewed by Contini et al., 2017), whereas other, behavioral, studies have suggested the opposite (Wu, Crouzet, Thorpe, & Fabre-Thorpe, 2015).

Journal of Cognitive Neuroscience 31:11, pp. 1699–1709 https://doi.org/10.1162/jocn_a_01444

Katholieke Universiteit Leuven

^{*}Present address: Center for Perceptual Systems, University of Texas at Austin

Furthermore, the presence of faces in the employed categories may hasten categorization (e.g., MEG: Cichy, Pantazis, & Oliva, 2014; monkey IT: Dehaqani et al., 2016), and thus the speed of categorization may depend on stimulus category. Furthermore, the speed of categorization may depend on whether the category is one preferred by the brain area where the signal originates, for example, faster for faces in face-selective areas than in object-selective regions. To have a handle on the category preference, we recorded from cortical regions that are activated more strongly and selectively by a particular category, that is, bodies (Sliwa & Freiwald, 2017; Premereur, Taubert, Janssen, Vogels, & Vanduffel, 2016; Lafer-Sousa & Conway, 2013; Popivanov, Jastorff, Vanduffel, & Vogels, 2012; Bell, Hadj-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). This departs from the previously cited studies of categorization dynamics, which recorded randomly within IT or had relatively poor spatial resolution (MEG).

Previously, we reported single-unit studies in two body patches that were identified with fMRI. One body patch is located in the lower bank of the middle STS and was labeled the mid STS body (MSB) patch. MSB is close to the face-selective patches middle lateral and middle fundus. The second body patch is located more anteriorly in the lower bank of the STS, close to face patch anterior lateral (AL), and was labeled the anterior STS body (ASB) patch. We found that, for both MSB and ASB neurons, the average response was stronger for visual images of bodies than for images of other categories (Kumar, Popivanov, & Vogels, 2019; Popivanov, Jastorff, Vanduffel, & Vogels, 2014). Single MSB and ASB neurons showed a marked within-category selectivity, responding only to some images of bodies, and some neurons even responded to images of objects, including faces (Kumar et al., 2019; Kalfas, Kumar, & Vogels, 2017; Popivanov, Schyns, & Vogels, 2016; Popivanov et al., 2014). Despite the degree of heterogeneity observed in the selectivity profiles at the singleunit level, the population responses of MSB and ASB neurons enabled a successful classification of images of bodies versus nonbodies (Kumar et al., 2019; Popivanov et al., 2014). Subsequent experiments showed that MSB and ASB neurons encoded body posture, body identity, and viewpoint, with ASB showing an encoding of the first two body parameters that was more tolerant to viewpoint transformations than MSB (Kumar et al., 2019). None of these studies, however, addressed the dynamics of the encoding of visual categories in these body patches, which is the topic of the present article.

An electrical microstimulation fMRI study suggested that ASB and MSB are anatomically connected (Premereur et al., 2016). Given the more anterior location of ASB, one would expect ASB to represent a higher processing stage than MSB. However, Kumar et al. (2019) failed to find any differences in response latencies between the two patches. Having recordings of the same stimulus sets in both patches of the same animals allowed us to compare the time courses and dynamics of encoding basic and superordinate categories in MSB and ASB.

METHODS

Participants

Two male rhesus monkeys (*Macaca mulatta*; 7–9 kg), the same animals as in Popivanov et al. (2014), were implanted with an MRI-compatible headpost and two recording chambers that targeted MSB and ASB, respectively. Animal care and experimental procedures complied with the Flemish, European, and National Institute of Health guidelines and were approved by the Animal Ethics Committee of KU Leuven.

Stimuli

Because the stimuli have been described before (Popivanov et al., 2012, 2014), we will provide only a brief description. We used 10 classes of achromatic images-monkey and human bodies (excluding the head), monkey and human faces, four-legged mammals, birds, two classes of human-made objects, fruits/vegetables, and body-like sculptures. Each class consisted of 10 images. The images of bodies depicted headless bodies in different postures and viewpoints. Similarly, the images of faces varied in viewpoint (profile to frontal views). To control for the difference in the aspect ratio of the monkey and human bodies, we presented two classes of human-made objects-one matching the aspect ratio of the monkey bodies (objectsM) and another one matching the aspect ratio of the human bodies (objectsH). We equalized the mean luminance and mean contrast across classes. The average object area per class was matched across classes, except for that of objectsH and human bodies, but allowing for some variation in area (range: 3.7° to 6.7° [square root of the area]) within each class. The mean vertical and horizontal extent of the images was 8.3° and 6.7°, respectively. The images were embedded in pink-noise backgrounds having the same mean luminance as the images and filled the display (height \times width: $30^{\circ} \times 40^{\circ}$). The stimuli were gamma corrected.

fMRI and Definition of MSB and ASB

Details of the fMRI procedure and analyses are provided in Popivanov et al. (2012). Briefly, monkeys were scanned using a block design during fixation of a small red target. Twenty images of six classes (monkey bodies, monkey faces, objectsM, four-legged mammals, birds, and fruits/ vegetables) were presented in blocks. The monkeys were scanned with a 3-T Siemens Trio scanner with an eight-channel coil (Ekstrom, Roelfsema, Arsenault, Bonmassar, & Vanduffel, 2008) and a gradient-echo single-shot EPI sequence (1.25-mm isotropic voxel size) after intravenous injection of the contrast agent monocrystalline iron oxide nanoparticle (MION; Feraheme, AMAG Pharmaceuticals Inc., 8–11 mg/kg). Activations for stimulus classes were computed with a general linear model.

We used the same criteria as those described by Kumar et al. (2019) to define the fMRI-based recording sites. Briefly, in both patches, voxels targeted for electrophysiological recordings showed significant activation for monkey bodies contrasted with objectsM. The MSB and ASB recording locations were verified using MRI scans with an MR-compatible guide tube with the electrode in the cortex during the MRI scan. In other confirmatory MRI scans, we visualized long glass capillaries, filled with copper sulfate or an electrode, that were inserted into the recording grid at selected positions. The recording locations were extrapolated from the trajectories of the imaged capillaries. The recording locations of the MSB and ASB recordings are described in Popivanov et al. (2014) and Kumar et al. (2019), respectively.

Electrophysiological Recordings

The electrophysiological procedures are described in Popivanov et al. (2014). In brief, single-unit recordings were performed with epoxylite-insulated tungsten microelectrodes (Frederik Haer Company). The electrode was lowered with a Narishige microdrive into the brain using a guide tube fixed in a Crist grid. After amplification and filtering, single units were isolated online using a custom amplitude- and time-based window discriminator. Eye position was measured with a video-based tracking system (SR Research EyeLink; sampling rate = 1 KHz). Stimuli were displayed on a CRT display (75-Hz vertical refresh rate) at a distance of 57 cm from the monkey's eyes. A digital-signal-processing-based computer system controlled stimulus presentation, event timing and juice delivery, and sampled eye position, spikes, behavioral events, and a photodiode-generated signal that indicated stimulus onset and end.

The stimuli were presented during passive fixation (fixation window size = $2^{\circ} \times 2^{\circ}$), as described in Popivanov et al. (2014). Stimuli were presented for 200 msec each with an ISI of approximately 400 msec. Fixation was required in a period from 100 msec prestimulus to 200 msec poststimulus to obtain a juice reward. In the pseudorandomization procedure, the 100 stimuli were presented randomly interleaved in blocks of 100 unaborted trials. Aborted presentations were not analyzed further. In the present analysis of the data of Popivanov et al. (2014) and Kumar et al. (2019), we included only neurons for which at least five unaborted presentations per stimulus were collected (MSB = 214)

neurons [Monkey E = 134], ASB = 146 neurons [Monkey E = 78]).

Data Analysis and Decoding

To select responsive neurons, we computed the firing rate for each unaborted stimulus presentation in 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. Responsiveness was tested using a split-plot ANOVA with Baseline versus response window firing rates as the repeated-measure factor and Stimulus as the between-trial factor. Only responsive neurons, defined by either a significant main effect of the repeated factor or a significant interaction between the two factors, were included in the decoding analyses.

The population decoding analyses were performed using the Neural Decoding Toolbox (Meyers, 2013). We used a support vector machine classifier with a linear kernel and the regularization parameter C equal to 1. Fivefold cross-validation was used where 80% of the trials were used for training and 20% were used for testing. Reported classification performances are for the independent test data. We used an "all-pairs" multiclass classification scheme (Meyers, 2013). The neural responses for each neuron were z score normalized using the mean and standard deviations of the training data per time bin.

We performed a set of decoding analyses, contrasting different category or stimulus labels. For each decoding analysis, we randomly sampled 68 neurons from MSB or ASB of each monkey and took the first five unaborted presentations of each stimulus to create pseudopopulation response vectors. We performed decoding for the data of each monkey separately and for the data pooled across the two monkeys. For each stimulus label, the responses for the five stimulus presentations were randomly concatenated to produce five pseudopopulation vectors (length = n neurons with a component being the response of neuron i on a given trial for the stimulus). We employed two types of decoding analyses. In the first type, same-stimulus classification (SSC), training and test vectors consisted of presentations of the same stimulus. For instance, for body versus nonbody SSC, we trained with four presentations of each of the body and nonbody stimuli and tested with the remaining presentation of each stimulus. Thus, during training, the classifier was exposed to responses to all exemplars (images) of a category. The performance of SSC is determined by (1) how well the classifier can separate the trialwise responses of the different categories and (2) the across-trial response variability of the responses to a stimulus. It examines the linear separability of the category representations of a body patch. The second type of classifier, different-stimulus classification (DSC), directly tested the generalization across stimuli of the same category. DSC was trained using all presentations (five trials per stimulus) of 80% of the stimuli of a category and tested with all presentations (five trials per stimulus) of the remaining 20% of the stimuli of the category. Thus, test and training vectors consisted of responses to different stimuli of a category. For instance, in the case of body versus nonbody DSC, responses to 80% of the body and nonbody images were used for training, and responses to the other images were used for testing. The performance of DSC depends on how well it generalizes to novel exemplars of the same category, a hallmark of categorization.

Each classifier was trained and tested using 20-msec bins that started 100 msec before stimulus onset and ended at 440 msec after stimulus onset, in steps of 20 msec. We ran the classifier for each bin 50 times with different resamplings of trials and neurons for construction of the pseudopopulation response vectors. The performance scores are the averages across resamplings. We trained and tested across all pairwise combinations of time bins, resulting in a temporal crosstraining (TCT; Meyers, 2013) matrix of classification performance scores.

The following category decodings were tested: (1) body (human and monkey bodies, four-legged mammals, and birds) versus nonbody (human and monkey faces, objectsM, and objectsH), (2) body (human and monkey) versus face (human and monkey), (3) body (human and monkey) versus object (objectsH and objectsM), (4) monkey body versus human body, and (5) monkey face versus human face. In addition, we decoded (1) the 10 monkey bodies, (2) the 10 human bodies, and (3) the 100 stimuli using SSC. For these exemplar decodings, we employed a step size of 10 msec to obtain a more accurate estimate of differences in the onset of stimulus encoding between MSB and ASB.

Statistical Testing

For each time bin, we randomly shuffled the (category) labels of the stimulus presentations and reran the SSC and DSC. This procedure was repeated 200 times, resulting in a null distribution of chance classification scores. With these null distributions, we computed, for each bin, the significance of the classification score for correctly labeled stimuli using false discovery rate correction for multiple comparisons (q < 0.005). The decoding onset latency was defined as the first 20-msec bin for which the decoding performance was significant. To test whether the decoding latencies of MSB and ASB differed, we generated a null distribution by performing the decoding analysis on 136 neurons randomly sampled (without replacement) 200 times from the pooled population of MSB and ASB neurons (272 neurons; 68 neurons from each monkey per region). We then assessed whether the decoding latency of the MSB sample was outside this null distribution (p < .005).

To test the significance of the differences between the TCT matrices of MSB and ASB, we generated 200 TCT matrices from 200 random samples (without replacement) of 136 neurons from the pooled MSB and ASB population. Next, we computed the pairwise difference between the classification scores for all possible pairs of the 200 TCT matrices, resulting in 19,900 difference matrices. We took the maximum and minimum values of each difference matrix, thus generating a distribution of maximum (positive) values and minimum (negative) values. The significance of the difference between TCT matrices of ASB and MSB (i.e., TCT matrix ASB - TCT matrix of MSB [excluding prestimulus onset bins]) was assessed by comparing each element of the difference matrix with the distributions of minimum and maximum values. The p values were computed as the percentage of data points of the null distribution falling above (maximum value distribution) or below (minimum value distribution) the observed difference. We plot classification difference scores in the difference matrices only for the matrix elements that had p < .01 (the magnitude of elements with p > .01 were set to zero).

RESULTS

We recorded the responses of 360 single units to 100 images of various categories in two fMRI-defined body patches, MSB and ASB, of two monkeys. The responses of these neurons, averaged across all stimuli, are shown in Figure 1C. Although ASB lies anterior to MSB, response-onset latencies of the population responses were similar. The population peristimulus time histogram (PSTH) of ASB showed a bimodal profile, having a dip in the response starting around 120 msec after stimulus onset. A similar dip was not apparent in the MSB population PSTH.

The population PSTHs, being averages across neurons and stimuli, cannot tell us whether differences exist in the evolution of object category encoding for the two patches. The selectivity, and not the response per se, is the critical factor relating the activity of body patch activity to perceptual behavior. To assess the temporal evolution of category selectivity for the MSB and ASB neuronal populations, we employed a decoding approach. We decoded categories at different levels, ranging from superordinate-level categorization (e.g., body vs. nonbody) to basic-level categorizations (e.g., human vs. monkey bodies). This allowed us to determine whether differences exist in the time courses of categorization at different levels. In addition, we compared classification performance for preferred and nonpreferred categories (e.g., human vs. monkey bodies compared with human vs. monkey faces). We employed two types of decoding, one (SSC) in which images of a category were identical for training and testing and a second one (DSC) in which the classifier was trained and tested with

Figure 1. The MSB and ASB patches: fMRI mapping and population spiking activity. (A) One example image from each of the 10 stimulus classes used to record the responses of the neurons in MSB and ASB. (B) Activations of the contrast "monkey bodies"-"objectsM" (t > 5). The images of Monkey B show an electrode targeting ASB. The figure is adapted from Kumar et al. (2019). (C) Population PSTHs of mean normalized firing rate averaged across the 10 stimulus classes for MSB (red; n = 214) and ASB (blue; n = 146) neurons. Data of the two animals were pooled. Shaded bands: + 1 SEM Normalization was performed for each neuron by dividing the response by the maximum mean firing rate (across the 10-msec bins). The gray-shaded region corresponds to the stimulus presentation (0 = stimulus onset).



different images within a category, which required generalization. Below, we will present the result of the decoding for the pooled data from both animals and briefly discuss the outcome of the decoding performed using the data of each monkey separately.

Both MSB and ASB neurons demonstrated excellent classification scores for the superordinate, body versus nonbody, classification (Figure 2A and B). As expected because of the more difficult generalization task, DSC performed less well than SSC in both patches. Although the peak accuracy occurred within the same time bin in both patches, MSB demonstrated a significantly (20 msec) faster categorization onset than ASB for both SSC and DSC. This same difference in categorization onset latencies was also present when decoding the data of each monkey separately. As for the mean responses (Figure 1C), the time course of the ASB performance accuracy was bimodal, with a dip between approximately 120 and 250 msec after stimulus onset (Figure 2A and B). This bimodal classification pattern was present in the data of each animal (data not shown), in contrast to that in MSB, which decreased monotonically after the peak.

The presence of the temporally bimodal classification pattern of ASB, absent in MSB, suggests a time-varying encoding of the category information in ASB. To examine the dynamics of these category encodings, we computed TCT matrices. The diagonal of the TCT matrix corresponds to the performance scores when training and testing time bins were identical (as in the first column of Figure 2A-F). The presence of high-performance scores concentrated along the diagonal would indicate a dynamic encoding; that is, the neural code is continuously evolving over time. Although the classification performance decreased moving away from the diagonal, the body-nonbody encoding was, to a large extent, consistent in MSB. Overall, MSB classification scores were well above chance when training and testing times differed. Note, however, that training at later bins generalized to the early test bins to a greater extent than training at an early bin generalized to later test bins. This asymmetry suggests differences between early and late category representations in MSB (see Discussion). The ASB neuronal population showed different encoding patterns at the beginning and in the later part of the response: The TCT plots showed two distinct squares. The second part of the response generalized only slightly to the first part of the response. The differences between the TCT matrices for MSB and ASB were statistically significant (difference matrices in Figure 2A and B) and were present for both DSC and SSC. The dynamic body versus nonbody categorization, with distinct encoding patterns for the first and second phases of the response, was present in ASB of each monkey.

Next, we asked whether similar differences between MSB and ASB might be present if the number of basiclevel categories entered into the decoding was restricted, thus decreasing the variability among exemplars of



Figure 2. Temporal patterns of body category decoding in MSB and ASB. (A, B) Left column: Time course of decoding accuracy for body versus nonbody for MSB (red) and ASB (blue) neurons for SSC (A, top) and DSC (B, bottom). Dashed lines indicate chance levels, and error bars correspond to standard deviations (n = 50 resamplings). The gray-shaded region corresponds to the stimulus presentation (0 = stimulus onset). The horizontal filled bars at the bottom indicate significant decoding accuracy for MSB (red) and ASB (blue). Second and third columns: TCT matrices describing the decoding performance (in color code; see the vertical bar on the right side of each panel) as a function of the training and testing time bins (width = 20 msec). Each row corresponds to the time bin that was used to train the classifier. The columns correspond to the time bins that were used to test the classifier that was trained using the data corresponding to the time bin indicated by a row. The diagonal corresponds to the decoding performance obtained when training and testing bins were identical. (A, top) TCT matrix for SSC for MSB (second column) and ASB (third column). (B, bottom) TCT matrix for DSC, same convention as in A (top). Right column: Differences between TCT matrices of ASB and MSB for SSC (A, top) and DSC (B, bottom). Only the significant classification difference scores in the difference matrices are shown, whereas the magnitude of difference matrix elements with p > .01 is set to zero. (C, D) The temporal patterns of body versus face category decoding for SSC (E) and DSC (D), organized as in A. (E, F) The temporal course of body versus object category decoding for SSC (E) and DSC (F), organized as in A.



Figure 3. Temporal patterns of basic-level category decoding in MSB and ASB. (A, B) Human versus monkey body category decoding: (A) SSC and (B) DSC. (C, D) Human versus monkey face category decoding in MSB and ASB: (C) SSC and (D) DSC. Same conventions as in Figure 2.

the decoded category. Decoding of body versus face (Figure 2C and D) produced results largely similar to those of body versus nonbody. First, decoding was about 20 msec faster in MSB than ASB for DSC. Second, the decoding profile was bimodal in ASB but not MSB. Third, the TCT plots showed dynamic decoding with two different decoding patterns in ASB but a greater consistency in MSB. Fourth, the decoding persisted for a longer period after stimulus exposure for ASB than for MSB. The generalization after 150 msec was stronger in ASB than that observed for the body versus nonbody categorization, which likely resulted from the overall higher categorization accuracy in this late part of the response for body versus face compared with body versus nonbody categorization. Similar trends were present for body versus object decoding (Figure 2E and F), although the bimodal

course of the decoding was less clear for ASB than for body versus face decoding, and the latency difference between MSB and ASB did not reach significance for DSC.

Next, we decoded basic-level categories from ASB and MSB data. We examined whether these patches differ with regard to the strength and time course of the encoding of basic-level categories and whether decoding onset latencies differ for superordinate and basic-level decoding. The first pair of basic-level categories we decoded was human versus monkey bodies. We were able to decode human versus monkey body with high performance scores from both patches, with little difference between the patches. The decoding was significantly faster in MSB compared with ASB, but the earlier onset proved to be significant only for the SSC. Interestingly, for both DSC and SSC, the classification persisted beyond stimulus exposure longer in ASB than in MSB (Figure 3A and B). The temporal course of the human versus monkey body categorization appeared to differ from that of the superordinate categorizations (e.g., body vs. nonbody) with a less clear bimodal time course in ASB. In fact, the TCT plots of the two patches were similar, except for a shift to longer latencies in ASB compared with MSB. The onset latencies of the categorization were the same for this basic-level classification and for superordinate classification in ASB. For MSB, the SSC onset latencies were identical for the different levels of categorization but tended to be 20 msec slower for DSC in the case of basic-level categorization.

As another basic-level categorization, we decoded human versus monkey faces, both of which are nonpreferred categories for these body patches. Not unexpectedly for such body patches, classification scores for faces were rather low, especially for the DSC (Figure 3C and D). The decoding onset appeared significantly later in ASB than in MSB, for both DSC and SSC, although the latency estimates were rather noisy because of the low classification scores. Furthermore, decoding was higher in ASB at later phases of the response, especially for DSC.

Finally, we decoded individual images, that is, decoding at the exemplar level. Figure 4A shows that individual stimulus decoding (100 images) was more accurate and faster in MSB compared with ASB. In addition, the TCT plots showed a strong decrease in the classification performance when training and test bins differed, although some generalization across time was present. Thus, the representation of individual images was rather dynamic in either ASB or MSB. There was also a tendency for a bimodal temporal profile in ASB but not in MSB. The decoding described above included exemplars from nonpreferred categories (objects and faces), which may have affected the decoding patterns. To address this, we also



Figure 4. Temporal patterns of exemplar decoding in MSB and ASB. Decoding of all individual images (A, N = 100) and the exemplars of the human body (B, N = 10) and the monkey body (C, N = 10) categories. Same conventions as in Figure 2. Only the SSC-type classification was possible for the exemplar decoding. No difference matrix between ASB and MSB is shown as there was a marked difference between the peak performances of MSB and ASB.

decoded individual images from the human and monkey body classes (10 exemplars each), both of which are preferred stimulus categories for the body patches. Overall, qualitatively similar results were obtained when individual stimuli from the human body (Figure 4B) and monkey body (Figure 4C) categories were decoded, although there was a trend toward higher generalization between early and later parts of the representations. Unlike decoding at the category level (see above), the decoding at the exemplar level is purely image based and may have resulted from differences in image size, body posture, viewpoint, or body identity (see Kumar et al., 2019). Because the interpretation of image-based exemplar decoding is difficult in the context of this study, we did not pursue this further and we refer the reader to Kumar et al. (2019), who systematically examined the decoding of viewpoint, posture, and identity of monkey bodies in body patches.

DISCUSSION

We examined the dynamics of category encoding in two cortical body patches using single-unit recordings examining both regions in each animal and using identical stimuli. For superordinate categorizations (body vs. nonbody, body vs. face, body vs. objects), the anterior body patch ASB showed a biphasic response pattern with different category representations in the early compared with the late response phase. The posterior body patch MSB showed earlier categorization onset and less dynamic encoding than ASB. The biphasic, dynamic encoding was less clear for the ordinate categorization of monkey versus human bodies, with ASB and MSB showing similar dynamics, except for a difference in average latency. We observed no consistent difference in the onset of ordinate (basic) level and superordinate level categorization where bodies were involved. The ordinate categorization of the nonpreferred categories of human versus monkey faces was relatively poor in both body patches. Qualitatively similar results were obtained whether training and test stimuli of the decoded categories were identical or not.

Previous human MEG studies (Grootswagers, Wardle, & Carlson, 2017; Cichy et al., 2014; Isik, Meyers, Leibo, & Poggio, 2014; Carlson, Tovar, Alink, & Kriegeskorte, 2013) and some macaque single-unit IT studies (Meyers et al., 2008) showed highly dynamic encoding of visual categories, with classification accuracy decreasing sharply at greater distances from the diagonals of the TCT plots. Here, we show less dynamic encoding than was observed in these studies, in agreement with other IT studies (Kumar et al., 2017; Zhang et al., 2011). The highly dynamic encoding observed in human MEG studies may result from the poor spatial resolution of this method, mixing signals of multiple areas that differ in both the nature and time courses of their category representations. This study shows that two patches in IT can already

show different dynamics for category decoding. Signals in MEG originate from a wider set of areas, which can result in considerable overestimation of the dynamic aspect of encoding in a particular area.

Using a linear decoder, we found that a small sample of MSB and ASB neurons can classify the superordinate body versus nonbody category with high accuracy. These results are consistent with our previous studies, which showed that the average MSB and ASB neural responses could distinguish between body and nonbody stimuli (Kumar et al., 2019; Popivanov et al., 2014). We observed a significant difference in decoding latency between the body patches, with MSB showing a categorization onset 20 msec earlier than ASB. This latency difference accords with the hierarchical organization of the IT cortex: Because of the more posterior location of MSB compared with ASB, one would expect that the relevant preferred-visual-category signals first appear in MSB. Kumar et al. (2019) reported no differences in response latency between ASB and MSB (see also Figure 1C). This apparent discrepancy between the present decoding results and those averaged population responses could be attributed to the higher sensitivity of the decoding analysis with regard to the timing of category representations in a population of neurons.

Our results provide evidence that body versus nonbody encoding was, to a considerable degree, static in MSB (but see below), whereas ASB showed two distinct neural activity signatures confined to the initial and late phases of the response. Similar biphasic dynamic decoding patterns were observed in ASB (but not MSB) for body versus face and body versus object categorization, indicating the robustness of these patterns. The origin of the biphasic decoding pattern in ASB is unclear. One possibility is that the dynamic encoding in ASB results from asynchronous inputs that carry information about different components of the stimulus at different times (Brincat & Connor, 2006). A second possibility is that local recurrent processing gives rise to the dynamic encoding observed in ASB. If true, this would imply that recurrent processing in MSB and ASB produces different representations at different periods during the response. Third, the distinct temporal encoding seen in the later phase of the response in ASB may represent feedback from other brain regions.

In the superordinate categorizations, MSB, in particular, demonstrated an asymmetry in decoding generalization over the course of the response: Training of the classifier at later bins generalized to the early test bins to a greater extent than training at earlier bins generalized to the later test bins. The existence of such temporal asymmetry in the generalization suggests that later category representations do differ from the early representations. This could be because of recurrent processing inside the body patch or input from other IT patches or from regions outside IT. However, because superordinate training at late-stage bins generalized well to early bins in MSB, there should still be considerable overlap between the early and late representations. Also noteworthy is that, for some categorizations (e.g., body vs. nonbody), decoding of the first 20 msec of the response generalized poorly to later bins, suggesting recurrent processing already present at the very early phase of the response. These interpretations and their implications for the neural encoding and temporal generalization require further investigation in future experimental and computational modeling studies.

MSB and ASB encoded the basic-level categories of monkey versus human bodies no earlier than the superordinate body versus nonbody category. This appears to conflict with the earlier decoding of basic-level categories compared with the superordinate animate-inanimate category reported for macaque IT by Dehaqani et al. (2016). The latter study performed random recordings within IT of neurons responding to various stimuli, whereas our data are from body patches that typically respond poorly to nonbody stimuli. We believe this to be a crucial difference: Our data show that superordinate categorization of body versus nonbody can be performed as rapidly as basic-level categorization of human versus monkey bodies by reading out a population of body patch neurons. The latency of categorization most likely depends on the category-selective properties of the neuronal population that is read out by the decoder. The longer latencies that we observed for basic-level categorization of the nonpreferred face stimuli are consistent with this idea. Another difference between our work and that of Dehagani et al. is that their superordinate category of animate exemplars contained both faces and bodies, which differs from the overall category selectivity of IT, in which faces and bodies form distinct categories (viz., face and body patches). Indeed, the classification accuracy for the animate-inanimate category in the Dehaqani et al. study was poor compared with the faster body versus face or primate versus nonprimate face categorization, and thus it might not be surprising that the highly abstract animate-inanimate distinction was decoded relatively late by the IT neurons they recorded. Note that other IT recording studies did not observe an animate-inanimate distinction (Yamins et al., 2014; Baldassi et al., 2013).

The main difference between ASB and MSB for the basic-level decoding was the later and longer-sustained decoding in ASB compared with MSB, but the overall dynamics were similar. The presence of a biphasic dynamic pattern in ASB for superordinate but less so for basic-level categories may suggest that the biphasic pattern is present for categories that require highly invariant processing. In this context, it is noteworthy that viewpoint-tolerant facial identity selectivity in the anterior face patch AM increases during the course of the response, reaching its peak relatively late in time (Freiwald & Tsao, 2010). Furthermore, further analysis of the face-patch data (Meyers, Borzello, Freiwald, & Tsao, 2015), using a decoding approach similar to this study, showed

a bimodal evolution of the classification of identityinvariant head poses in AL and AM, but not the posterior face patch, ML. These findings imply that stimulus representations of the anterior face patches change during the course of the response, perhaps similarly to what we have observed here for body categorization in ASB.

Our observation of differences in the category encoding dynamics of the two IT body patches suggests that caution is necessary when interpreting decoding data derived from IT neurons that have been recorded from dispersed, random locations within the extensive IT cortex (Majaj, Hong, Solomon, & DiCarlo, 2015). Decoding performance depends on which neurons are read out by the decoder, and efferent projections can differ between IT regions (Kravitz, Saleem, Baker, Ungerleider, & Mishkin, 2013). This anatomy suggests that a single decoder that weights inputs arising from the whole of IT, as is assumed when decoding from a collection of randomly located regions spread across IT, is biologically unrealistic. Understanding how the rich stimulus selectivity observed in IT patches is translated into behavior requires knowing how the outputs of IT neurons are combined in areas that control behavior. This will need refined causal perturbation methods (Jazaveri & Afraz, 2017) and combined registrations in IT and its output regions. A recent study of V1-V2 connectivity (Semedo, Zandvakili, Machens, Yu, & Kohn, 2019), which suggested that only particular V1 population activity patterns ("communication subspace") affect V2 responses, hints at the complexity of this sort of readout.

Acknowledgments

We thank 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. S. Raiguel for critical reading of the draft. The MSB data are from recordings performed by Dr. I. Popivanov. The fMRI mapping data were obtained by Dr. I. Popivanov, in collaboration with Dr. J. Jastorff and Dr. W. Vanduffel. This work was supported by the Fonds voor Wetenschappelijk Onderzoek Vlaanderen (grants G.0932.14N and G.00007.12-Odysseus).

Reprint requests should be sent to Rufin Vogels, Neurofysiologie, Campus Gasthuisberg, Herestraat 3000, Belgium, or via e-mail: rufin.vogels@kuleuven.be.

REFERENCES

- Baldassi, C., Alemi-Neissi, A., Pagan, M., Dicarlo, J. J., Zecchina, R., & Zoccolan, D. (2013). Shape similarity, better than semantic membership, accounts for the structure of visual object representations in a population of monkey inferotemporal neurons. *PLoS Computational Biology*, 9, e1003167.
- 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.

Brincat, S. L., & Connor, C. E. (2006). Dynamic shape synthesis in posterior inferotemporal cortex. *Neuron*, *49*, 17–24.

Carlson, T., Tovar, D. A., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics of object vision: The first 1000 ms. *Journal of Vision*, 13, 1.

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience*, *17*, 455–462.

Contini, E. W., Wardle, S. G., & Carlson, T. A. (2017). Decoding the time-course of object recognition in the human brain: From visual features to categorical decisions. *Neuropsychologia*, 105, 165–176.

Dehaqani, M. R., Vahabie, A. H., Kiani, R., Ahmadabadi, M. N., Araabi, B. N., & Esteky, H. (2016). Temporal dynamics of visual category representation in the macaque inferior temporal cortex. *Journal of Neurophysiology*, *116*, 587–601.

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.

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.

Freiwald, W. A., & Tsao, D. Y. (2010). Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science*, *330*, 845–851.

Grootswagers, T., Wardle, S. G., & Carlson, T. A. (2017). Decoding dynamic brain patterns from evoked responses: A tutorial on multivariate pattern analysis applied to time series neuroimaging data. *Journal of Cognitive Neuroscience*, *29*, 677–697.

Isik, L., Meyers, E. M., Leibo, J. Z., & Poggio, T. (2014). The dynamics of invariant object recognition in the human visual system. *Journal of Neurophysiology*, 111, 91–102.

Jazayeri, M., & Afraz, A. (2017). Navigating the neural space in search of the neural code. *Neuron*, *93*, 1003–1014.

Kalfas, I., Kumar, S., & Vogels, R. (2017). Shape selectivity of middle superior temporal sulcus body patch neurons. *eNeuro*, 4, ENEURO.0113-17.2017.

Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., & Mishkin, M. (2013). The ventral visual pathway: An expanded neural framework for the processing of object quality. *Trends in Cognitive Sciences*, *17*, 26–49.

Kumar, S., Kaposvari, P., & Vogels, R. (2017). Encoding of predictable and unpredictable stimuli by inferior temporal cortical neurons. *Journal of Cognitive Neuroscience*, 29, 1445–1454.

Kumar, S., Popivanov, I. D., & Vogels, R. (2019). Transformation of visual representations across ventral stream body-selective patches. *Cerebral 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. *Nature Neuroscience*, *16*, 1870–1878.

Majaj, N. J., Hong, H., Solomon, E. A., & DiCarlo, J. J. (2015). Simple learned weighted sums of inferior temporal neuronal firing rates accurately predict human core object recognition performance. *Journal of Neuroscience*, *35*, 13402–13418. Meyers, E. M. (2013). The neural decoding toolbox. Frontiers in Neuroinformatics, 7, 8.

Meyers, E. M., Borzello, M., Freiwald, W. A., & Tsao, D. (2015). Intelligent information loss: The coding of facial identity, head pose, and non-face information in the macaque face patch system. *Journal of Neuroscience*, *35*, 7069–7081.

Meyers, E. M., Freedman, D. J., Kreiman, G., Miller, E. K., & Poggio, T. (2008). Dynamic population coding of category information in inferior temporal and prefrontal cortex. *Journal of Neurophysiology*, *100*, 1407–1419.

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.

Popivanov, I. D., Schyns, P. G., & Vogels, R. (2016). Stimulus features coded by single neurons of a macaque body category selective patch. *Proceedings of the National Academy of Sciences, U.S.A.*, 113, E2450–E2459.

Premereur, E., Taubert, J., Janssen, P., Vogels, R., & Vanduffel, W. (2016). Effective connectivity reveals largely independent parallel networks of face and body patches. *Current Biology*, *26*, 3269–3279.

Semedo, J. D., Zandvakili, A., Machens, C. K., Yu, B. M., & Kohn, A. (2019). Cortical areas interact through a communication subspace. *Neuron*, 102, 249–259.

Sliwa, J., & Freiwald, W. A. (2017). A dedicated network for social interaction processing in the primate brain. *Science*, *356*, 745–749.

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.

Wu, C.-T., Crouzet, S. M., Thorpe, S. J., & Fabre-Thorpe, M. (2015). At 120 msec you can spot the animal but you don't yet know it's a dog. *Journal of Cognitive Neuroscience*, 27, 141–149.

Yamins, D. L., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences, U.S.A.*, 111, 8619–8624.

Zhang, Y., Meyers, E. M., Bichot, N. P., Serre, T., Poggio, T. A., & Desimone, R. (2011). Object decoding with attention in inferior temporal cortex. *Proceedings of the National Academy of Sciences, U.S.A.*, 108, 8850–8855.