🔓 OPEN ACCESS

Intracortical recordings reveal the neuronal selectivity for bodies and body parts in the human visual cortex

Jesus Garcia Ramirez[a,b,1], Michael Vanhoyland[a,b,c,1] 🆔, N. A. Ratan Murty[d,e,f], Thomas Decramer[a,c] 🆔, Wim Van Paesschen[g] 🆔, Stefania Bracci[h] 🆔, Hans Op de Beeck[i], Nancy Kanwisher[d,e,f] 🆔, Peter Janssen[b,2] 🆔, and Tom Theys[a,c]

Affiliations are included on p. 11.

Body perception plays a fundamental role in social cognition. Yet, the neural mechanisms underlying this process in humans remain elusive given the spatiotemporal constraints of functional imaging. Here, we present intracortical recordings of single- and multiunit spiking activity in two epilepsy surgery patients in or near the extrastriate body area, a critical region for body perception. Our recordings revealed a strong preference for human bodies over a large range of control stimuli. Notably, body selectivity was driven by a distinct selectivity for body parts. The observed body selectivity generalized to non-photographic depictions of bodies including silhouettes and stick figures. Overall, our study provides unique neural data that bridge the gap between human neuroimaging and macaque electrophysiology studies, laying a solid foundation for computational models of human body processing.

body perception | electrophysiology | visual cortex | human | intracortical

Perceiving bodies is a fundamental component of social vision in humans and animals. This complex process engages a large specialized cortical network (1), with the Extrastriate Body Area (EBA) as its cornerstone in humans (2). Despite its importance, a critical gap exists: no study has recorded the responses of neurons to images of bodies in the human visual cortex. This leaves the actual neural computations that underlie the visual perception of bodies in humans poorly understood. In contrast, single-unit data in nonhuman primates [NHPs (2)] provide rich insights into the neural mechanisms of body processing, but similar evidence in humans is currently lacking (3).

This study aims to tackle a number of specific questions about the neural representation of visually presented bodies in humans. Are individual neurons in the human visual cortex (4) selective for whole bodies, or body parts, and if the latter, are they selective for specific body parts (like hands)? Do body-selective responses arise very soon after stimulus presentation, as expected for a largely feed-forward process? How do we process more abstract depictions of bodies, such as outline drawings or stick figures? Are neural populations selectively responsive to photographs of bodies also selectively responsive to abstract representations? These questions can only be addressed using intracortical recordings of body-selective neurons in the human visual cortex. We were fortunate to record single- and multiunit spiking activity from intracortical electrode arrays implanted in or near the EBA in two neurosurgical patients undergoing invasive epilepsy monitoring.

The neuronal populations in this region exhibited a clear and rapid preference for human bodies over other visual stimuli, including faces, objects, and other body categories. Subsequently, we performed a series of experiments to comprehensively characterize the pattern of body-selective responses in the recorded arrays. A clear preference for human bodies was found, as well as selectivity for individual body parts. At the individual site level, the majority of body-selective sites coded for multiple body parts simultaneously, with distinct preferences for each body part. The body-selective representations within the recorded arrays extended well beyond photographic depictions of bodies, as predicted by previous fMRI work (4). Overall, our findings provide the most direct evidence to date for body selectivity in the human extrastriate visual cortex, and start to reveal the neural representations and computations underlying this important component of social perception.

Results

We performed microelectrode recordings (96-electrode Utah array) in the lateral occipitotemporal cortex (LOTC) of two patients with refractory epilepsy undergoing intracranial recordings for seizure detection (Fig. 1*A*). CT/MRI coregistration revealed that

Significance

Our ability to perceive and understand the world relies heavily on the visual cues provided by the human body, which are processed in the extrastriate body area (EBA). Here, we provide comprehensive characterization of the tuning properties of neurons in or near the human EBA. We found a strong preference for human bodies over a range of control stimuli, driven by a distinct selectivity for body parts. The neural representation of bodies exhibited partial tolerance to image rotation and generalized to silhouettes of bodies and stick figures. Our study reveals the neuronal selectivity, tolerance to image transformations, latencies, and mesoscale organization in a region of the human visual cortex which has never been investigated at the neuronal level.

[1]J.G.R. and M.V. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: peter.janssen@kuleuven.be.
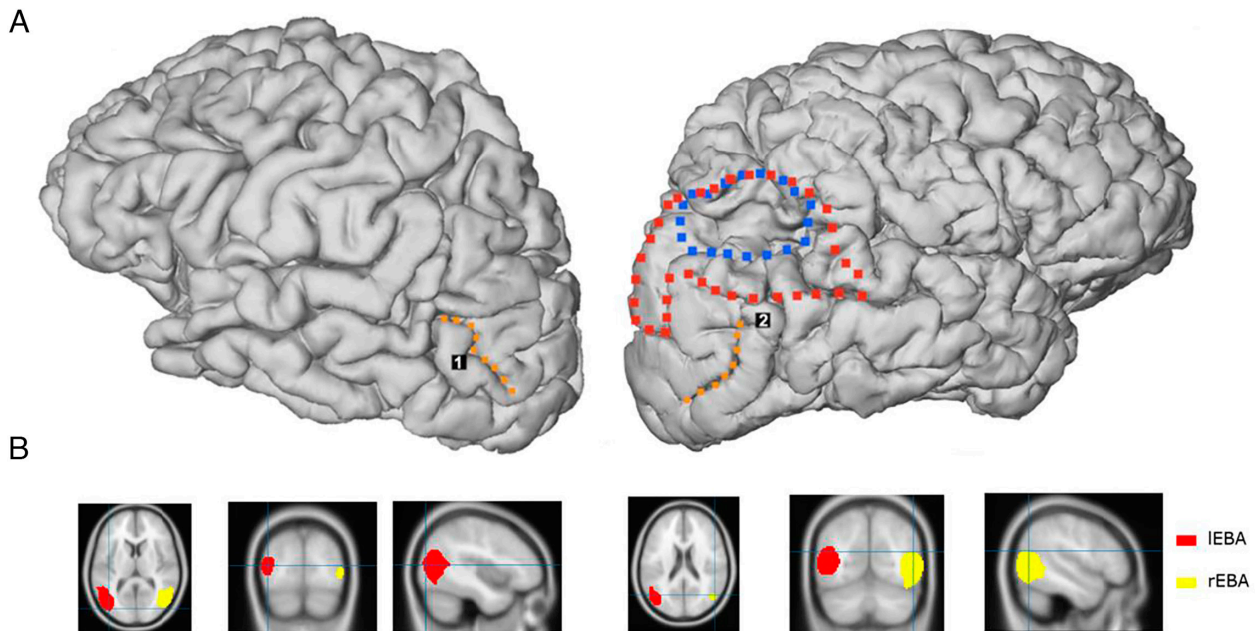
Fig. 1. Multielectrode array recording sites. (*A*) Anatomical Utah array location on 3D rendering of the patient's brain for *Left* hemisphere of P1 (MNI coordinates: X = −41, Y = −83, Z = 9) and *Right* hemisphere of P2 (MNI coordinates: X = 51, Y = −66, Z = 19). Note the previous resection cavity (blue dotted line) above the array in P2 and the presumed epileptogenic zone (red dotted line) that was resected during electrode removal. Dotted orange lines represent the Lateral Occipital Sulcus (LOS). (*B*) Location of the Utah arrays relative to across-subjects EBA parcels identified in Julian et al. (5). Axial, coronal, and sagittal view for P1 (*Left*) and P2 (*Right*). Crosshair indicates the position of the Utah array for each patient.

both arrays were implanted in the lateral occipital cortex at MNI coordinates consistent with the EBA (Fig. 1*B*).

Category Selectivity

To determine the visual responsiveness and underlying selectivity of the neural population, we first recorded the responses to 100 stimuli containing images of bodies, faces, and other object categories [stimuli from Popivanov et al. (6), Fig. 2*A* for examples]. We obtained data from 128 visually responsive MUA sites across two subjects (two sessions in array P1 and one session in array P2, yielding $n_{S1}^{P1} = 59, n_{S2}^{P1} = 46, n^{P2} = 23$ visually responsive sites, respectively).

The response profile for an example recording site is shown in Fig. 2*B*. The multiunit activity (MUA) of this site responded significantly more to images of (human and animal) bodies than to images of (human and animal) faces or objects (one-way Kruskal–Wallis ANOVA, $P = 6\text{e}{-}24$, Fig. 2 *B*, *Upper* panel). For the majority of body images, the responses were well above the highest response to any face or object image (Fig. 2 *B*, *Lower* panel).

The population response selectivity was relatively heterogeneous (Fig. 2*C*). On average, we observed a marked preference for images of bodies over other stimulus categories (see average responses in *Lower* panel of Fig. 2*C*). To quantify category selectivity for each individual site, we computed a category-selectivity index d' (7, 8) (*Methods*). Significance was assessed by comparing the category d' index to a null distribution generated by randomly permuting category labels across trials. More than half of the recording sites across the population were indeed body-selective (N = 67/128 body-selective sites, permutation test, $P < 0.05$; 57.1% and 30.4% for array P1 and array P2 respectively, χ^2 test, $P < 4.17\text{e}{-}11$, $\chi^2 = 43.53$). We also detected a smaller number of sites selective to faces (N = 27/128, permutation tests, $P < 0.05$, arrows in Fig. 2*B*), and only one MUA site was object selective (N = 1/128, permutation test threshold, $P < 0.05$). Collectively, the observed body-selective responses and the anatomical location of the electrode recording sites in the lateral occipital cortex indicate that the recording arrays were located in or close to the EBA (4).

Next, we aimed to characterize the strength and time course of the population selectivity for human bodies in the recorded arrays. We employed linear decoders on the responses recorded in each session to assess the extent to which the pattern of neural activity elicited by a human body differed from that elicited by other categories. Decoding was performed in a time-resolved fashion on a trial-by-trial basis (*Methods* for details). Specifically, we tested the neural pattern discriminability between human bodies and 1) the aggregate of the other categories (body vs nonbody) and 2) the images within the body category (i.e., human bodies, monkey bodies, animals, and birds).

We found that a linear decoder trained on the population response could robustly discriminate between a human body and the other categories on a trial-by-trial basis in every recording session (permutation tests for each session $P < 0.005$; Fig. 2*D*). Decoding performance was nearly optimal (AUC close to 1) for both recording sessions in P1, while the decoding in array P2 was less accurate, presumably due to noise and the lower number of body-selective sites. The time courses of body selectivity were comparable across arrays in terms of the onset of selectivity (between 100 and 150 ms after stimulus onset) and the time of peak performance (250 ms after stimulus onset). This finding is consistent with the only other prior human invasive study examining body-selective responses around the EBA (3) using subdural grid recordings (electrocorticography or ECoG). This study reported body-selective local field potential responses peaking at around 260 ms after stimulus onset in a grid macrocontact contact at the approximate location of the EBA.

The recorded populations could reliably discriminate between different body subcategories with high accuracy, suggesting the presence of subordinate category representations (permutation tests $P < 0.01$; *SI Appendix*, Fig. S1*A*). The confusion matrices associated with the subcategory-decoding (*SI Appendix*, Fig. S1*B*) show that human bodies were the easiest to distinguish from other
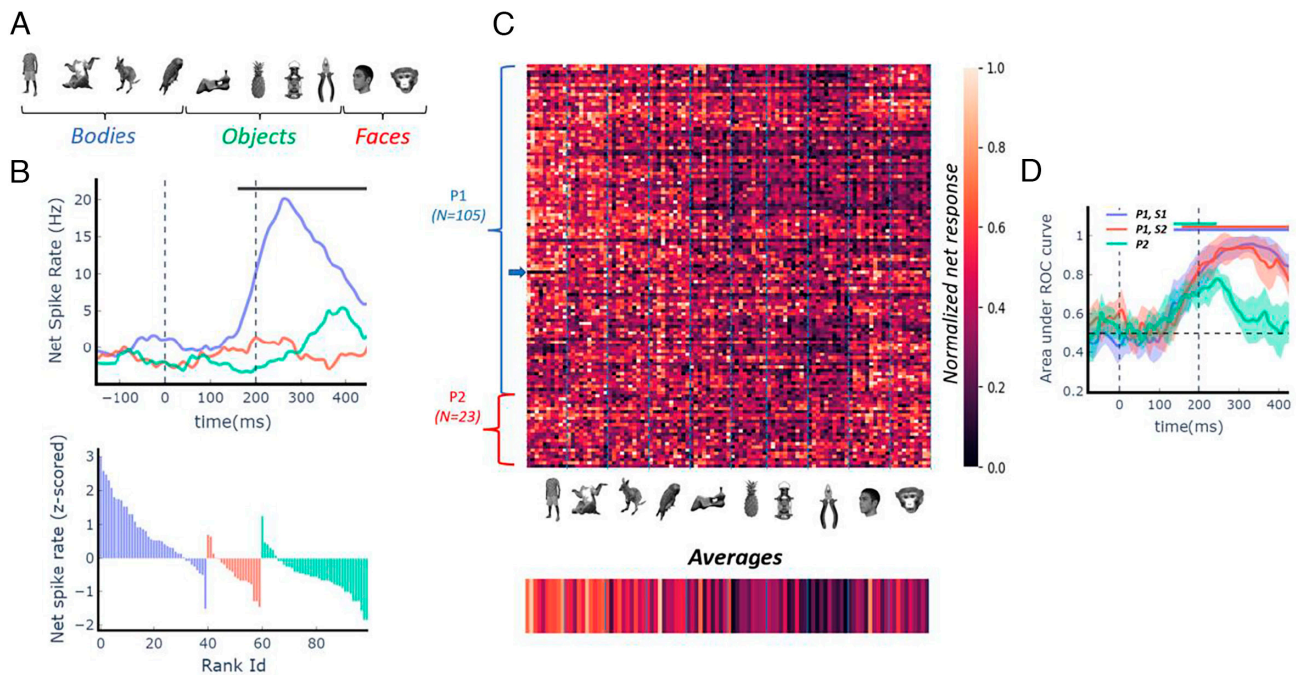
Fig. 2. Category selectivity. (*A*) Example stimuli from the category selectivity experiment grouped into Bodies (Blue), Objects (Green), and Faces (Red). (*B*) Responses of a representative multiunit example site. *Top* panel shows the time course of the mean net spike rate per category group. Horizontal bar represents significant response differences between categories (uncorrected ANOVA test per 100 ms time point, sliding window of 10 ms, $P < 0.05$). *Bottom* panel, average net z-scored responses to individual stimuli ranked within their respective category group (bodies, objects, or faces). (*C*) Color-coded normalized net responses of all visually responsive sites in array P1 (105 sites) and array P2 (23 sites) for the presented stimuli. Each row represents one recording site and each column represents one image. Columns were sorted by image category. (*D*) Time course of decoder performance (area under the ROC curve; AU-ROC) for each recording session (blue: session 1 in P1; red: session 2 in P1; green: P2). Lines and shading show mean ± 1std classification AU-ROC for decoders trained to differentiate between stimuli containing a human body versus all other stimuli. Distribution reflects 10-fold grouped cross-validation. Horizontal bars indicate above-chance classification ($P < 0.005$; one-sided permutation test with at least five consecutive significant time points).

subcategories. To confirm these findings, we trained individual decoders to discriminate between each body subcategory in a one-vs-all fashion and found that decoders trained to discriminate human bodies versus all other body subcategories exhibited the best decoding performance in every recording session (*SI Appendix,* Fig. S1C). To visualize the neural population representation, we performed a multidimensional scaling (MDS) analysis on the pairwise neuronal distances at the window of peak body selectivity (*Methods*). The MDS-derived two-dimensional neuronal object space (*SI Appendix,* Fig. S2) further demonstrates that images of human bodies form a separate cluster distinct from the images of most other tested categories. Overall, our findings are in line with previous human studies that demonstrated a preference for human bodies in body-selective visual areas (9).

It is possible that all the results we report thus far were driven solely by the selectivity for low-level visual features. To address this concern, we compared the decoding results derived from our recorded neural data with those obtained using a state-of-the-art computational model of V1 activity [VOneBlock (10), *Methods*]. For both arrays (P1 and P2), human bodies were significantly easier to discriminate from the other categories using the recorded neural data compared to the V1 model (*SI Appendix,* Fig. S3A, permutation test threshold, $P < 0.001$ for both sessions of array P1, $P < 0.05$ for array P2), suggesting that the neural selectivity for human bodies cannot be solely accounted for by low-level visual feature tuning. When further comparing the decoding performance for human bodies against the rest of the body subcategory images, the results remained significant for array P1 (*SI Appendix,* Fig. S3B, permutation test threshold, $P < 0.05$ and $P < 0.001$ for Session 1 and Session 2 of array P1 respectively), indicating a robust selectivity for human bodies over other body types. However, for array P2, while within-category discrimination

was higher for the recorded neural data compared to the V1 model, but this difference did not reach statistical significance (*SI Appendix,* Fig. S1D, $P = 0.18$).

Body Selectivity Is Driven by Features Prevalent in Human Bodies

Our previous findings showed an overall preference for bodies over other tested stimulus categories. Next, we explored how the recorded populations responded to specific body parts. To this end, we presented a stimulus set that included images of whole bodies, isolated body parts (hands, feet, bodies without hands and feet, and heads), and images of objects (examples in Fig. 3A).

We recorded 85 visually responsive MUA sites in both subjects (one session in array P1 and one session in array P2, yielding $n^{P1} = 57, n^{P2} = 28$ visually responsive sites, respectively). To quantify if the recorded neural population encoded body part-specific information, we trained linear decoders to 1) distinguish between individual body parts and nonbody images (e.g., feet versus objects; "body part decoders") and 2) discriminate between different body parts ("intrabody decoder").

The body part decoders robustly discriminated between every individual body part and images of objects in both patients (permutation tests for each session $P < 0.005$; Fig. 3B). The time course of decoding varied slightly for each body part and reached peak performance around 250 ms after stimulus onset for array P1 and around 200 ms for array P2, triangles in Fig. 3C. Note that this time window also corresponds to the time at which human body selectivity emerged in the previous analysis (Fig. 2D). We found near-optimal performance (area under ROC curve at perfect 1.0) for each individual body part in array P1, and a somewhat lower performance in array P2, in which the body-part information was
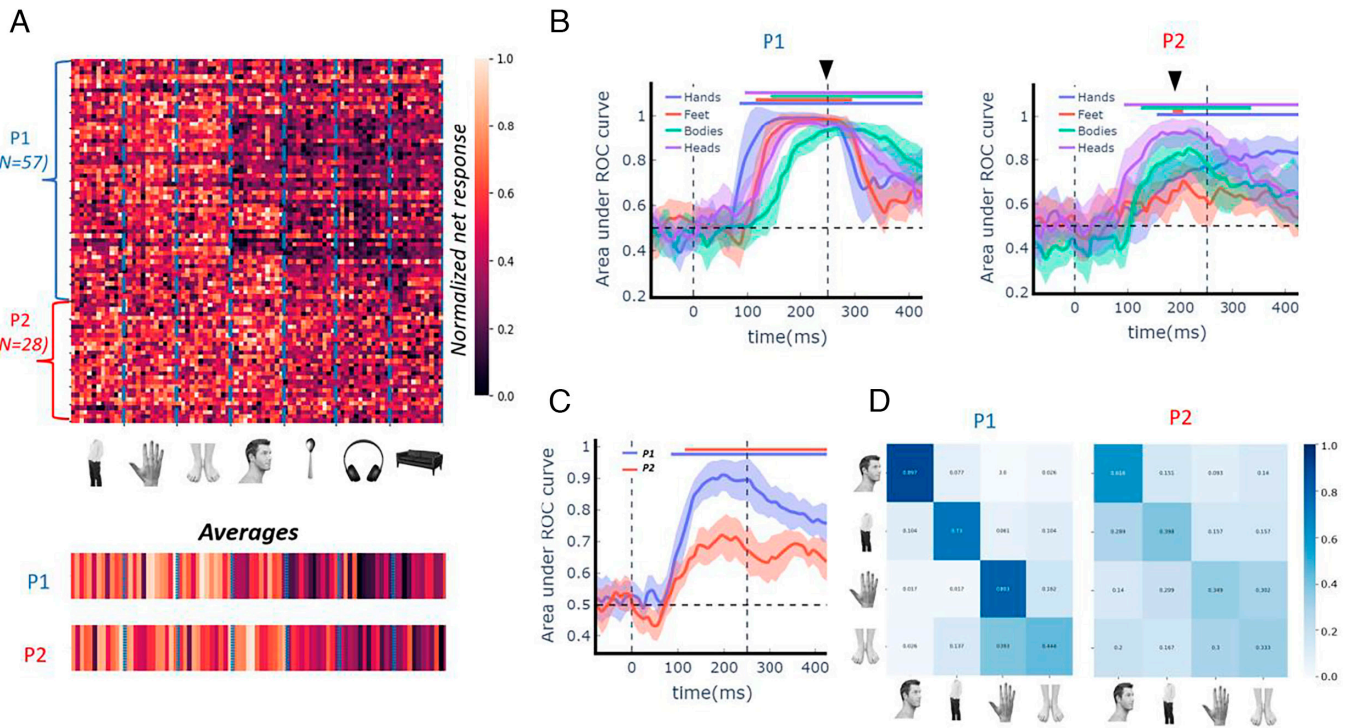
Fig. 3. Body part selectivity. (*A*) *Top* panel depicts the normalized responses of all visually responsive sites in P1 (blue, 57 sites) and P2 (red, 28 sites) to 84 images of whole bodies, isolated body parts (hands, feet, bodies without hands and feet, and heads), and images of objects (12 images per category, example stimuli shown in *Middle* panel). Each row represents one site and each column represents one image. *Bottom* panel shows the average response to each image across all visually responsive sites in array P1 and array P2. (*B*) Time course of performance of individual body part decoders (AU-ROC) for array P1 (*Left*) and array P2 (*Right*) trained to distinguish between individual body part and nonbody images. Same conventions as in Fig. 2*D*. (*C*) Time course of performance of intrabody part decoders (AU-ROC) for both arrays trained to discriminate between different body parts. Same conventions as in Fig. 2*D*. (*D*) Confusion matrices for intrabody decoders at the peak decoding time window for array P1 (*Left*) and array P2 (*Right*). The x-axis represents ground-truth labels and the y-axis the predicted labels. The matrices are normalized across rows to highlight the proportion of true positives across each category.

dominated by a preference for heads and bodies (i.e., torsos with arms and legs) over hands and feet. Furthermore, the recorded neural population could reliably discriminate between different body parts (permutation tests for each session $P < 0.005$; Fig. 3*C*) suggesting fine-grained body-part representations at the population level. This pattern was also more prominent in array P1 than in array P2 (Fig. 3*D*). The MDS-derived two-dimensional neuronal object space (*SI Appendix*, Fig. S4) further confirmed our results since the images containing a body part (whole bodies, heads, hands, feet) were separable from images containing an object.

We also tested these results against a CNN-based V1 model and found (*SI Appendix*, Fig. S5) that the fine-grained body-part representations emerged exclusively with our recorded data. For array P1, the body-part decoders trained using neural data significantly outperformed the V1 model across all body parts, while for array P2, all body parts except feet showed significantly better decoding, consistent with a preference for bodies and heads (*SI Appendix*, Fig. S5*A*). Intrabody decoding also revealed superior performance for both patients compared to the V1 model (*SI Appendix*, Fig. S5*B*). Taken together, these findings suggest that the body-selective section of the lateral occipital complex encodes specific body-part information that is linearly separable from nonbody images.

Finally, we explored the representation of each body part in individual body-selective sites. We asked whether these sites were "selective" or "general" in their responses to different body parts. A selective site was defined as a site that had a significant d' value (compared to objects) for only one specific body part (example selective single unit in *SI Appendix*, Fig. S6), while a general site was defined as one that had a significant d' value for more than one body part (example general single unit in *SI Appendix*, Fig. S7). The recorded populations consisted predominantly of

general sites for both arrays (91.4% and 85% for array P1 and array P2 respectively, *SI Appendix*, Fig. S8*A*). Such a result however could be a consequence of multiple single units selective for different body parts recorded by the same electrode. To address this concern, we identified pure single units (SUA) using the Plexon offline spike sorter (48 units in array P1 and 34 units in array P2), and found that 94.1% of body-selective neurons in array P1 and 92.9% in array P2 were general in their responses.

To further investigate how these general sites encoded different body parts, we examined the spatial distribution of the selectivity for each body part (hands, feet, bodies without hands and feet, and heads), with respect to objects, across both arrays (*SI Appendix*, Fig. S8*B*). In array P1, we observed two distinct neural populations: one pool of highly selective sites for heads (left side of array P1) and another pool that responded to other body parts with different preferences. The results for array P2 indicate primarily body-selective sites with a preference for heads and bodies. Collectively, these results point toward a more diffuse code where body-selective sites encode multiple body parts simultaneously.

Tolerance to Shape-Preserving Image Transformations

To test the tolerance of the recorded neural representations to shape-preserving image transformations, we recorded responses to 1) silhouette versions of the category-bodies stimuli and to 2) bodies at different orientations.

Silhouettes lack internal details, which allow to study the degree to which neural response selectivity depends on the external contours alone. We reasoned that if the body contour is a strong determinant in driving the responses of body selective neurons,

intact images of bodies and body silhouettes would share a common representation. We therefore tested whether decoders trained on intact body images (discriminating between images of human bodies and nonbodies) generalized to silhouette versions of the same images without retraining (Fig. 4A, *Methods*). We found that in both arrays, the trained body decoders could generalize from natural to silhouette versions of bodies, and vice versa. The decoding performance and observed time course for silhouettes were almost identical to the original decoders trained on intact images (Fig. 4B). Furthermore, in array P1, the generalization performance could not be attributed to tuning for low-level V1-like visual features (Fig. 4C, permutation values, $P < 0.001$). Note that while array P2 exhibited higher generalization performance for neural data compared to the V1 model, this difference was not statistically significant (Fig. 4C, $P = 0.11$). Individual sites did also generalize, representing intact and silhouette bodies similarly (*SI Appendix*, Fig. S9). These results indicate a shared neural code for intact and silhouette body stimuli.

To explore tolerance to changes in body orientation, we computed the average correlation between each population vector at a reference orientation and the population vectors corresponding to the same stimulus at a different orientation (*Methods*). First, we examined the reliability of the recorded populations for each orientation separately (by computing the average correlation of each population vector at a particular orientation in half of the trials with the population vector in the other half of the trials). Significant information was present for all orientations except one in array P1 (permutation tests $P < 0.05$, Fig. 4 D, *Left* panel), while array P2 only showed significant information for two orientations (permutation tests $P < 0.05$, Fig. 4 D, *Right* panel). This suggests that neurons in array P2 showed limited tolerance to in-plane changes in orientation, as the available information was

limited to a small number of orientations (45° and 215°) that were not spatially contiguous.

Next, we focused on array P1 to examine the similarity of the neural representations across orientations. The population of body-selective sites in array P1 showed tolerance to body orientation over a range of approximately 45° (Fig. 4E), yet the population representation changed markedly for larger orientation changes. Overall, our findings indicate that the responses of body-selective sites to images of bodies depend greatly on their orientation, highlighting the importance of orientation in shaping the neural representation of bodies.
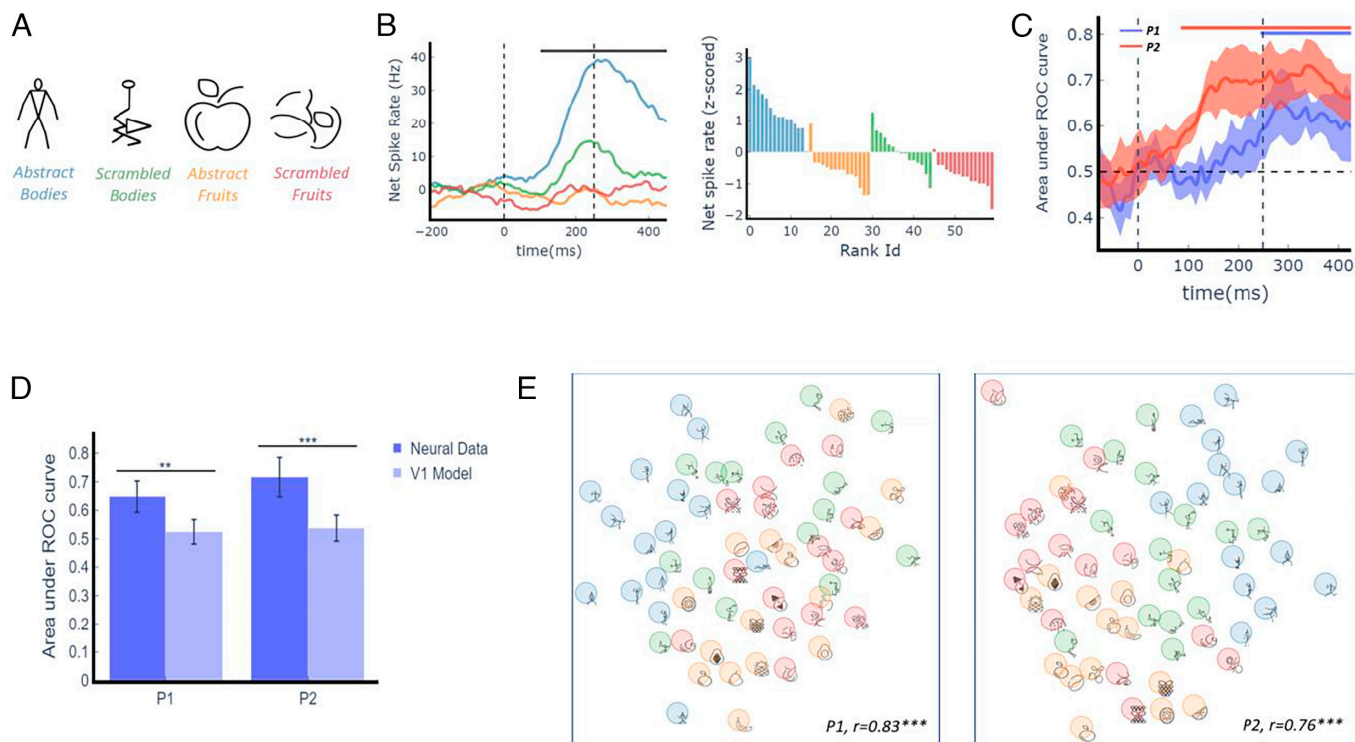
Representation of Abstract Bodies

Finally, we investigated how body-selective neural sites encode highly abstract images of bodies. Stick figures are highly impoverished compared to naturalistic images of bodies, but humans have no problem correctly identifying and classifying these images as bodies. We compared the responses of stick figures with scrambled stick figures (to account for low-level visual features) together with outline drawings of objects (abstract fruits) and scrambled objects (example stimuli shown in Fig. 5A). We recorded responses from 35 visually responsive sites in both subjects (one session in array P1 and one session in array P2, yielding $n^{P1} = 13$, $n^{P2} = 22$ visually responsive sites, respectively). The example site in Fig. 5B responded significantly more (Kruskal–Wallis ANOVA, $P < 5e{-}22$) to images of stick bodies than to scrambled controls or fruits (Fig. 5 B, *Left* panel) and, on average, the large majority of stick bodies evoked stronger responses than scrambled bodies (Fig. 5 B, *Right* panel).

At the population level, linear decoders trained to distinguish among the four stimulus classes performed significantly above
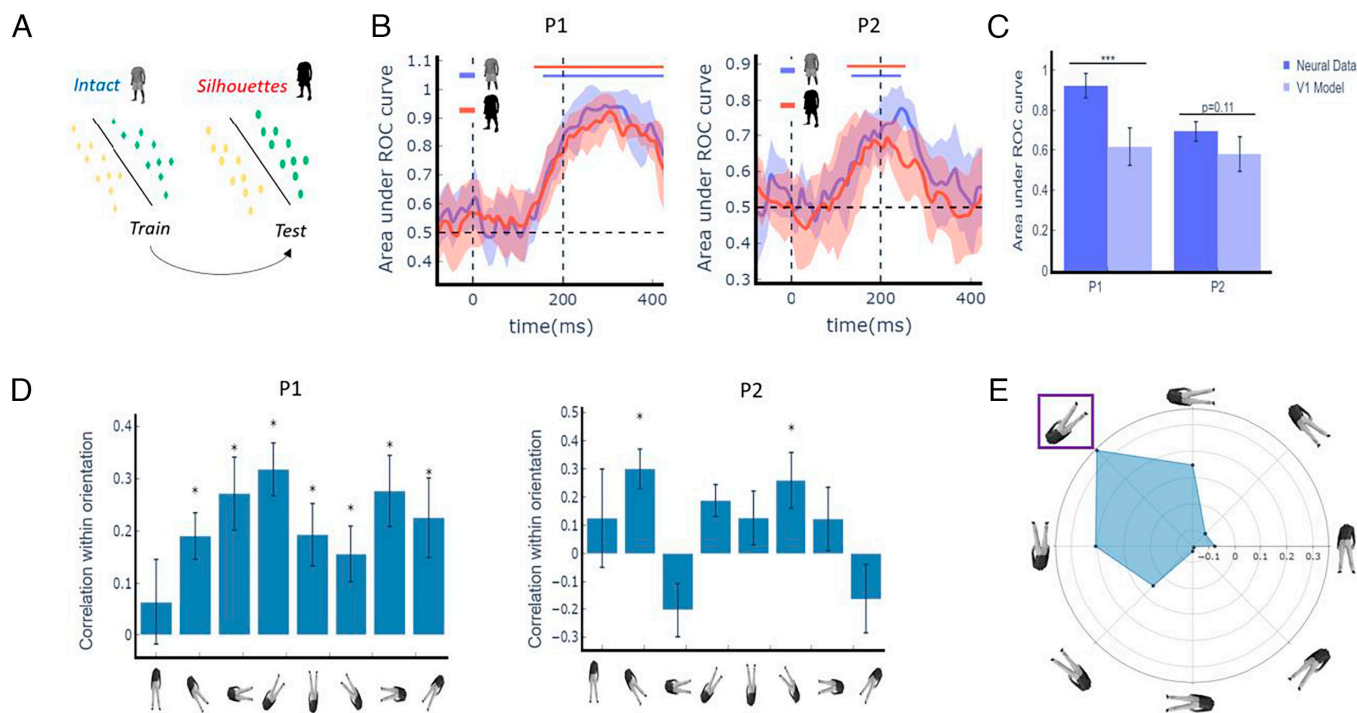
Fig. 4. Tolerance to shape-preserving image transformations. (A) Schematic illustration of decoders employed to investigate the shared population representation between intact images of bodies and their corresponding silhouettes. Body decoders were trained on intact images to discriminate between images of human bodies and nonbodies and tested on silhouette versions of the same images. (B) Time course of body decoders' performance (AU-ROC) for array P1 (*Left*) and array P2 (*Right*), for intact images (blue), and generalization decoders for silhouettes (red). Same conventions as in Fig. 2D. (C) Comparison of peak performance for generalization decoders trained with neural data compared to a model of V1-like activity. (D) Average correlation of each population vector with itself (autocorrelation) across independent halves at each orientation for array P1 (*Left*) and array P2 (*Right*). (E) Tolerance to body orientation in array P1. The polar plot illustrates the average correlation between each population vector at the orientation with the highest autocorrelation (purple square) and the population vectors corresponding to the same stimulus at different orientations. *$P < 0.05$, ***$P < 0.001$, one-sided permutation test.

Fig. 5. Responses to abstract bodies. (*A*) Example stimuli from the different categories presented to assess abstract representation of bodies. (*B*) Representative example site. *Left* panel, time course of mean net spike rate for abstract bodies (blue), scrambled bodies (green), abstract fruits (orange), and scrambled fruits (red). Horizontal bar represents significant response difference between categories (uncorrected ANOVA test per time point, *P* < 0.05). *Right* panel: averaged z-scored net responses to individual stimuli and ranked within their respective category. (*C*) Time course of linear decoders performance (AU-ROC) for both arrays (P1 in blue and P2 in red). Lines and shading show mean ± 1std classification AU-ROC for each decoder trained to discriminate the category of each stimulus. Same conventions as in Fig. 2*D*. (*D*) Comparison of peak performance for decoders trained with neural data (*C*) compared to a model of V1-like activity. (*E*) Visualization of similarity relations in the population evoked response for the presented stimuli obtained using multidimensional scaling (MDS) for P1 (*Left*) and P2 (*Right*). r corresponds to the Pearson correlation between projected distances in 2d (MDS-transformed) and the actual neural distances in the n-dimensional space, where n corresponds to the number of visually responsive sites. **P < 0.01, ***P < 0.001, one-sided uncorrected permutation test.

chance level for both patients (corrected permutation tests P < 0.005, Fig. 5*C*). This performance cannot be explained by tuning for low-level visual features (Fig. 5*C*, permutation test, P < 0.01 and P < 0.001 for arrays P1 and P2, respectively). The confusion matrices (*SI Appendix*, Fig. S10) show that stick bodies were more discriminable than other stimulus categories. This is also depicted in the associated MDS (Fig. 5*D*) where the intact stick figure cluster (blue) is clearly distinct from their scrambled counterparts (green) and intact and scrambled fruit-like images.

These findings suggest that specific neural populations in LOC respond preferentially to bodies even when the images are reduced to stick figures. This result however does not address the question of whether the neural representation of stick bodies overlaps with the neural representation of full-cue presentations of human bodies. To explore this further, we presented a reduced version of the original EBA set localizer of Downing et al. (4), which contains different bodies and body parts together with stick figures of bodies (only one session in array P2, yielding 32 visually responsive sites; *Methods*). We first confirmed our previous findings by showing that every individual body category was linearly separable from the other object categories (corrected permutation tests P < 0.005, *SI Appendix*, Fig. S11*A*), including images of bodies when presented as stick figures (corrected permutation tests P < 0.005, *SI Appendix*, Fig. S11*B*). Next, we performed a decoding generalization analysis analogous to the silhouette experiment. Specifically, we tested whether linear decoders trained to discriminate between images of bodies and objects could generalize when tested with stick figures (intact versus scrambled). We observed partial but significant generalization as the ability of the classifier to generalize

in this way reached significance but was considerably reduced compared to the original classifier (*SI Appendix*, Fig. S11*C*). Moreover, generalization performance was significantly higher with the recorded neural data compared to V1-like activity (*SI Appendix*, Fig. S12*B*, permutation test, P < 0.01). These results suggest partially overlapping population representations of human bodies and stick figure bodies. Consistent with this idea, the two-dimensional MDS neural representation showed separate clusters for bodies and objects, with stick figure bodies placed in between (*SI Appendix*, Fig. S11*E*).

Discussion

We present direct neural recordings in a body-selective patch in the human visual cortex. Through a series of experiments, we characterized body-selective responses within the lateral occipital cortex at the level of single- and multiunit activity. Our results show that the neural responses in the human body selective cortex exhibit a clear preference for human bodies over other visual stimuli tested (including faces, objects, and other body categories), are driven by individual body parts, show orientation dependence, and generalize from naturalistic (photographic) images of human bodies to highly symbolic/abstract images of bodies.

Intracortical recordings in the human visual cortex are still exceptional but represent a highly valuable addition to the repertoire of research methodologies in systems neuroscience. Unlike fMRI or other imaging techniques, utah array recordings of spiking activity allow to unambiguously determine the properties of individual human neurons and small populations of neurons, their

response and selectivity latencies, tolerance to stimulus transformations and the submilimeter organization of neuronal properties in a 4 by 4 mm region of the cortex. Thus, the spatial and temporal resolution of the recorded body selective responses is unprecedented in humans and sheds light on the neural mechanisms supporting the visual perception of bodies.

Because we could not acquire fMRI data prior to array insertion, the relation between the fMRI-defined EBA of each individual patient and our recording sites remains uncertain, but the observed pattern of body-selectivity and the implantation sites of the electrode arrays in the lateral occipital cortex strongly suggest that the Utah arrays recorded in or near the EBA (4), a body selective region broadly implicated in human body perception (9, 11). Note also that the location of the fMRI-defined EBA (as illustrated in Fig. 1) is based on group averaging across a large number of subjects, while in a given individual the EBA may be much smaller or scattered over several voxels. Therefore, an additional and crucial advantage of our utah array recordings is that they were derived from the individual brain rather than from group-based averaging. While body-selectivity has been studied extensively using functional MRI, this is direct evidence for body selectivity in humans at the level of multiunit spiking activity [Pourtois et al. (3) for local field potential responses on a single macroelectrode contact of a subdural grid]. Few prior fMRI studies (12, 13) also suggested neural sensitivity for body parts in the human EBA, but since fMRI measures the pooled activity of multiple thousands of cortical neurons, it remained unclear whether selectivity for individual body parts would be observed at the neuronal level in humans.

Our study provides empirical support for the existence of a body selective neural population tuned to discriminate between individual human body parts. This preferential coding for human body parts compared to nonhuman parts aligns with similar findings in the human fusiform cortex regarding human versus nonhuman faces (14). We also observed nontrivial differences in the preferences for different body parts across arrays. In particular, array P1 (located more ventrally) had more decodable information about each body part [with higher responses overall for hands and feet, which agrees with previous fMRI studies (15, 16)], while the more dorsally located array P2 showed a clear preference for heads and bodies (without head, hands, and feet). Orlov et al. (17) previously used reverse correlation analyses to suggest a mosaic of regions (encompassing the EBA) that respond differently in subtle but systematic ways to different body parts. Qualitative comparisons suggest that array P1 likely recorded from a region previously reported as the "upper-limbs" region (which mostly overlaps with the EBA), while array P2 may have been located in the "upper-face" area. We also observed that body selective responses in array P1 emerged faster and were stronger compared to array P2. Both patients also differed in their clinical history (P2 previously had a resection close to the array location), but the striking correspondence between the neuronal properties and previous fMRI studies suggests that the between-patient differences were mainly related to the anatomical locations of the arrays (see also Supplementary information for anatomical scans postimplantation). Despite these differences, all main findings of this study (body and body part selectivity, contour selectivity, orientation tolerance, representation of stick figures) were present in both patients.

We found that body-selective sites encode multiple body parts simultaneously, with distinct preferences for each body part. This finding eliminates the possibility of single units encoding semantic labels such as "finger," "hand," or "foot," as the neurons we recorded responded to different body parts. Instead, the strong generalization from intact bodies to silhouettes suggests that the body representations depend on the similarity between shape features, which aligns with the role of shape features in driving responses in the ventral occipito-temporal cortex (18, 19). However, the features that drive the recorded populations are more complex than simple low-level features. This becomes evident when we compare our decoding analyses with a state-of-the-art computational model emulating V1 activity. Most likely, the lower decoding performance in array P2 resulted from a lower signal to noise ratio in array P2 recordings as well as fewer body-selective channels (*SI Appendix*, Fig. S13). Both factors are known to lower the ability to decode information from the neuronal population. Critically, all the decoding trends in array P2 were in line with the findings in array P1.

Downing et al. (4) previously showed that the body-selective representations within the EBA may generalize to other abstractions such as silhouettes, line drawings, and stick figures. It is crucial to confirm such findings with direct measurements of local neuronal spiking activity. In the available literature using noninvasive methods, seemingly abstract representations could have been an artifact of i) nearby yet separate representations (not visible in imaging due to limited spatial resolution), ii) input from other regions through feedback (not visible due to limited temporal resolution), or iii) below-threshold input from other regions (not visible unless method separates action potentials from field potentials). Our findings demonstrate some generalization for all abstraction types even when measuring action potential output directly at high spatial and temporal resolution. We show that the degree of generalization at the level of individual neurons differs across abstraction types. We found strong generalization from intact bodies to silhouettes, but the neural representation of stick figures exhibited only partial alignment with the neural representation of photorealistic bodies. Crucially, this partial generalization vanished with scrambled counterparts, indicating that low-level features alone cannot explain stick figure responses.

Body selectivity has also been observed in the macaque extrastriate cortex (6, 20). The macaque mid-STS body (MSB) patch has been suggested as the homolog of the human EBA, even though neural properties had never been formally compared. Our findings reveal a striking similarity between the properties of human body-selective neurons and those found in the MSB (21). Notably, we observed similar selectivity for intact and silhouette bodies consistent with Popivanov et al. (22), and the recorded neurons exhibited a limited tolerance to image rotation, with strong responses to specific body parts, as in Popivanov et al. (6). Population-level analyses demonstrated accurate decoding of human versus monkey bodies; in addition, faces or heads were distinguishable from bodies and body parts, in line with Kumar and Vogels (23).

Our study also finds that response latencies of body-selective neurons in humans were slightly longer (>100 ms) than those reported from MSB in monkeys (70 ms) (8). Determining response latencies is difficult and depends on many factors (number of trials, luminance and contrast of the stimulus, antiepileptic medication) which are different in our recordings compared to experiments in macaque monkeys. Moreover, in a previous study (24) the range of response latencies to images of objects (85 to 195 ms) was consistent with macaque inferotemporal data. Another possible explanation for this difference might be the cortical expansion of human brains (about 14 times larger) (25) compared to macaques. Note that our observed latencies align with the only other prior invasive human study examining body-selective responses around the EBA (3) using subdural grid recordings (electrocorticography or ECoG). The response latencies reported in our study provide a rare insight into the timescale of information processing in the human visual cortex, a dimension not captured by conventional human fMRI

methods. These findings provide crucial evidence for the macaque monkey as a valid model for studying body processing in the visual system. Future studies will have to include intracortical recordings in other nodes of the human body processing network to chart the homology between human and macaque body processing systems. Unfortunately, such recordings in humans can only be performed on exceedingly rare occasions.

With a body and a face as integral parts of human anatomy, it is natural to wonder about the ways in which these two elements interacted in body selective neurons. In the category selectivity experiment, we presented well-isolated faces (Fig. 2*A* for example images) and found that approximately 25% (N = 27/128) of the visually responsive sites responded selectively to faces. This finding aligns with previous studies in the macaque extrastriate cortex, where some neurons in the MSB have been shown to respond well to faces in addition to bodies (8, 26). The presence of face-selective neurons within the body selective cortex corroborates earlier findings demonstrating that face-selective neurons can be found outside face patches (27). This heterogeneity in neuronal selectivity at the single site level could not be detected in fMRI studies because the fMRI signal is the result of averaging across a large population of individual neurons with different response properties.

Further insights into the neural underpinnings of body perception will require an end-to-end mechanistic understanding of the specific visual features that drive the selectivity of body selective neurons (28). One possible approach is the use of computational models. Deep neural networks have been used to successfully predict the responses of single neurons in macaques (29–32) and fMRI bold activations in humans (33, 34), including category-selective regions like the EBA (35, 36). Our high-resolution intracortical recordings from the human EBA can be used not only as a strong experimental constraint for computational models, but also as a tool to uncover the specific visual features that drive responses of human cortical neurons [for example using artificially synthesized stimuli, Ratan Murty et al. (35)]. There are several ways to extend our work using such approaches. First, our data offer the possibility of developing a model per recording site, which can be compared against previous fMRI-based pooled models of the entire EBA. Second, our data provide ground-truth responses to images that fall well outside the manifold of naturalistic images used in previous studies (such as silhouettes and abstract bodies). Whether or not predictions from naturalistic image generalize to stimuli outside the training distribution remains an open question (37, 38). Third, electrophysiological data in our study are dynamic and the temporal resolution of MUA recordings provides an additional constraint for (generally feedforward) models (39). Together, the synergy between computational approaches and high-resolution spiking neuronal activity from human EBA promises to uncover the neural mechanisms of body perception in humans at a finer-grained and quantitative level.

Methods

Study Approval. We obtained ethical approval (study number s53126) for conducting microelectrode recordings using the Utah array in patients with epilepsy. The study protocol (s53126) was approved by the local ethical committee (Ethische Commissie Onderzoek UZ/KU Leuven). The study was carried out in accordance with the principles of the Declaration of Helsinki, the principles of good clinical practice, and all applicable regulatory requirements set by the Federal Agency for Medicines and Health Products (FAGG).

To ensure safety and compliance, strict adherence to imposed safety measures was maintained. This included the use of case report forms and detailed reports on any (serious) adverse events. All data collected during the study were encrypted and stored at the University Hospitals Leuven. The study was discussed with the

patients during the preoperative consultation, which took place more than six weeks before the surgery. Patients were informed that the goal of the study was to gain more insight into the network involved in epilepsy ("Microscale Dynamics of Epileptic Networks: Insights from Multiunit Activity analysis in neurosurgical patients with refractory epilepsy," Bougou et al., EANS 2023, Barcelona) and to understand brain functions better through the use of microelectrodes. They were also informed that the insertion of the microelectrode array involves a slightly increased risk of infection and local hemorrhage. Furthermore, patients were assured that they could opt out of the study at any time. Written informed consent was obtained from the patients on the evening prior to their surgery.

Clinical Information. We obtained invasive intracranial recordings from the lateral occipital cortex in 2 refractory epilepsy patients. Patient 1 was a 58-y-old woman who had never undergone brain surgery. Patient 2 was a 29-y-old male who had undergone resective epilepsy surgery in the right temporoparietal cortex at the age of 19 [MNI coordinates anterior border (64, −42, −32), superior border (51, −53, 56), posterior border (55, −67, 42), and inferior border (60, −60, 27)]. The pathology report after this resection mentioned chronic meningitis. After a year of seizure freedom, his seizures recurred and invasive intracranial recordings were performed as diagnostic epilepsy surgery workup. Patient 1 was on Levetiracetam 2,000 mg, Diphantoine 100 mg, Rivotril 1 mg, and Lacosamide 300 mg, all taken twice daily. Patient 2 was on Levetiracetam 2,000 mg and Lacosamide 300 mg twice daily. During hospitalization, medication was adjusted in both patients according to the number of detected seizures.

Utah arrays were located adjacent to the clinical macroelectrodes, consistent with previous studies using microelectrode arrays (40–45). Specifically, they were located in the left lateral occipital cortex (MNI coordinates −41, −83, 9) in patient 1 and right lateral occipital cortex in patient 2 (MNI coordinates 51, −66, 19). Since the target locations of intracranial electrodes were solely determined by the epileptologist and based on electroclinical findings and noninvasive multimodal imaging, our study protocol did not allow to acquire fMRI data prior to implantation. The multielectrode arrays were implanted via the craniotomy performed for the implantation of subdural grids and were thus inserted in close proximity to this grid. No additional incisions were made specifically for the purpose of this study. Arrays were inserted in or near the presumed epileptogenic zone (based on preoperative multimodal imaging) to study the microscale dynamics of the epileptic network. Consequently, the implantation site was considered a potential resection area prior to recording. Invasive recordings revealed no clear ictal onset in patient 1. In patient 2, seizures originated from the inferior border of the previous resection, and a new resection was performed at the time of electrode removal. In both patients, the array was outside the actual epileptogenic zone. Importantly, none of the patients has experienced complications related to the microelectrode array. After 14 d the arrays were removed together with the other clinical intracranial electrodes in a second surgery.

Postelectrode removal, both patients underwent MRI scans to assess potential implantation-related complications. The postoperative CT scan (with electrodes in situ) was coregistered with the postoperative MRI (after electrode removal) using Brainlab© Elements software to evaluate any structural changes resulting from electrode insertion (*SI Appendix*, Fig. S16). Postoperative MRI (T1, T2, 3D FLAIR) revealed no structural alterations such as gliosis, ischemia, or hemosiderin deposition at the implantation site. Additionally, clinical neurological examinations conducted after electrode removal indicated no functional deficits. In a previous study, we demonstrated the safety of Utah array implantations and even observed intact functional MRI activations postarray removal, providing evidence of preserved brain tissue functionality at the implantation site (40).

Invasive Recordings. We recorded from a 96-electrode array (Utah ArrayTM – Blackrock Neurotech) from day 1 until day 14 after array implantation. We inserted the array using a pneumatic inserter wand provided by Blackrock Neurotech. We closed the dura above the array and placed the bone flap on top to secure the array. We placed reference wires subdurally and ground wires epidurally. We digitally amplified the signal using a Cereplex M head stage, connected to a 128-channel neural signal processor (Blackrock Neurotech), and sampled the signal using Central Software at a sampling rate of 30 kHz. We applied a 750 Hz high-pass filter for spiking activity. We set the multiunit detection threshold at 3 SD of the noise band. We considered units recorded on the same electrode but on different days as separate units. The variable signal quality in the immediate

postoperative period resulted in a difference in the number of responsive and selective sites during each recording session. All spike sorting was performed offline (Offline Sorter 4, Plexon, TX).

Stimulus Presentation. We conducted experiments in a dimmed hospital room. We utilized custom-built software on a 60 Hz DELL-P2418HZM LED monitor to present stimuli. Patients sat 60 cm away from the screen (1 pixel = 0.026). We instructed them to focus on a small red square (0.2 × 0.2°) positioned at the center of the display. We continuously monitored the position of their left or right pupil using a dedicated eye tracker (Eyelink 1,000 Plus, 1,000 Hz) in head free-to-move mode. If they deviated from the electronically defined 3° by 3° fixation window, the trial was aborted. To synchronize the data, we attached a photodiode to the upper left corner of the screen, which detected a bright flash coinciding with the first frame of the stimulus. This flash was invisible to the patient. To accurately record baseline spike rate activity, we introduced an intertrial interval between the offset of the stimulus and the onset of fixation for the next trial. Additionally, to maintain fixation, we instructed the patients to press a button with their right hand whenever a distractor (red or green cross) randomly appeared at the fixation point in approximately 2% of the trials.

Stimuli.

Category-bodies. This stimulus set consisted of 100 achromatic stimuli grouped over 10 categories (6), each stimulus was presented ~7 to 10 times. Categories included mammals, birds, fruits, bodies (human–monkey), faces (human–monkey), objects (matched to the aspect ratio of images from the human class "objectsH" and monkey class "objectsM", respectively), and sculptures. The mean luminance values were equated across classes. The mean vertical and horizontal extent of the images was 6.0° and 4.6°. Stimuli were presented on a uniform gray background. After a fixed fixation period of 300 ms, stimuli were presented for 200 ms with an intertrial interval of 150 ms.

Silhouette tolerance. We presented silhouette versions of every stimulus from the category-bodies set randomly interleaved with stimuli from the categorical set in one session for each patient. Each stimulus was presented ~7 to 10 times. The size of the silhouette version of every stimulus corresponded to their intact counterpart. Stimuli were presented on a uniform gray background.

Body part responses. We presented a stimulus set that included images of bodies and isolated body parts as well as images of objects. The images from every category except for the feet categories are from the set employed by Matic et al., 2020 (46). The included conditions from this set comprised three categories of different body parts: bodies (only torso; without the head, hands, and feet), hands, and heads and three object categories validated with behavioral judgments (i.e., tools, manipulable objects, and nonmanipulable objects). The stimuli from the different object categories were assembled such that category information could not be extracted from their physical shape. All images were controlled for low-level visual properties (e.g., spatial frequencies and luminance) by means of the SHINE toolbox (47). Each condition consisted of 12 achromatic images (~12 repetitions per stimulus) on a white background (Fig. 3A). The mean vertical and horizontal extent of the images was 8.3° and 8.1°.

Tolerance in-plane body rotations. Four bodies from the category selectivity task were shown within the receptive field at 8 different rotations (45° angles between each rotation, starting from the upright position). Body size was 4°. For patient 1, stimulus location was (at 6°, 3.5°) in the lower contralateral quadrant, for patient 2 stimuli were presented foveally (at 0°, 0°). Each stimulus was repeated ~12 times for every rotation. Patients performed a distractor task as described above.

Abstract body representations. To investigate how body-selective neurons respond to abstract representations of human bodies, we utilized stick figures as abstract representations, while also employing scrambled versions of these stick figures as a control to account for low-level visual features. Stimuli consisted of 60 black line drawings on a white background grouped in 4 categories (bodies, scrambled bodies, fruits, and scrambled fruits). Each body and fruit image was matched to 1 scrambled stimulus which contained the same line fragments. Scrambled counterparts were manually made by reorganizing the individual lines composing the original body or fruit. The mean vertical and horizontal extent of the images was 6.6° and 6.0°. Patients performed a distractor task as described above (~2% of trials).

Reduced EBA localizer. This stimulus set consisted of 125 achromatic stimuli grouped over 6 categories; each image was presented ~7 to 10 times. Categories included whole bodies (photographic and drawings), body parts (photographic and drawings), objects (photographic and drawings), places, stick figure bodies, and scrambled stick figures. The images from each category, with the exception of places, come from the original EBA localizer of Downing et al. (4). Stimuli were presented on a uniform white background. The mean vertical and horizontal extent of the images was 6.8° and 6.3° of visual angle. Patients performed a distractor task as described above (~2% of trials). Fixation duration was 300 ms, stimulus duration 250 ms, and intertrial interval 150 ms.

Imaging. We acquired a T1 weighted image using a 3-T MR scanner (Achieva dstream, Philips Medical Systems, Best, the Netherlands) at the University Hospitals Leuven for presurgical planning and to reconstruct the postoperative electrode location. The imaging parameters were as follows: time of echo 3.1 ms, inversion time 900 ms, slice thickness 1.0 mm, 400 × 400 matrix, and 283 coronal slices. Following electrode placement, a CT scan (Siemens) was performed with a slice thickness of 1 mm, a voltage of 120 kV, and a dose length product of 819 mGy cm. The CT scan served to verify the electrode locations and to rule out any hemorrhage.

For imaging preprocessing, we utilized SPM12 software (Wellcome Department of Cognitive Neurology, London, UK) running on MATLAB (Mathworks, Natick, MA). The preprocessing steps included (1) Setting the anterior commissure as the origin for both the MRI and CT scan (2) Realigning the images (3) Coregistering MRI and CT images and (4) Warping the MRI and coregistered CT to a common high-resolution brain atlas [MNI1525 (48)]. To generate cortical 3D renderings and for the high-resolution MNI atlas, we employed Freesurfer. The precise electrode location on the cortical surface and the MNI coordinates were determined using iElectrodes Software (49).

Single-Site Data Analysis. We computed multiunit firing rates for each correct trial in two analysis windows: a baseline window ranging from 150 to 50 ms before stimulus onset and a response window ranging from 200 to 300 ms after stimulus onset. We restricted all analyses to visually responsive sites. To assess the responsiveness of each recorded site, we performed independent t-tests between the spike rates in the baseline and response windows for each stimulus category separately (e.g., baseline vs human bodies). For each individual experiment, we considered a site as visually responsive if any of these comparisons was significant ($P < 0.01$). The reasoning behind this criterion is that sites with a strong preference for a single category could have been considered as not responsive when testing baseline vs response across all stimuli in the test. All subsequent analyses were based on the baseline subtracted, average net firing rates. In cases where responses across sites were shown on the same figure, we scaled the average net firing rates for each site in the range ($x_{min} = 0$, $x_{max} = 1$) as follows:

$$r(i) = \frac{r(i) - r_{min}}{r_{max} - r_{min}}.$$

In which, $r(i)$ corresponds to the average net response for stimulus i for that site r_{max} and r_{min} correspond to the "best" and "worst" stimuli for that site, respectively.

Calculation of d'. To quantify category selectivity for individual sites, we computed d prime index or d' (7, 8), which takes into account the mean responses to the corresponding categories as well as the variability of the responses to the individual images of a category. The d' indices were computed for different pairs of categories using the net evoked firing rates. Here, we illustrate the computation of d' in the example case of comparing trials in which an image of a body or nonbody was presented:

$$d'(body) = \frac{\mu_{body} - \mu_{non-body}}{\sqrt{\frac{\sigma^2_{body} + \sigma^2_{non-body}}{2}}}.$$

In which μ_{body} and $\mu_{non-body}$ are the mean firing rates and σ^2_{body} and $\sigma^2_{non-body}$ are the variance in firing rates on trials in which an image of a body or nonbody was presented, respectively. For this analysis, we removed the responses to images from the birds and mammals categories as they included both a body and a face. The significance of each site's d' was assessed by comparing the obtained values to a null distribution of values generated by randomly permuting category labels (body or nonbody) across trials (1,000 iterations).

Noise-Ceiling Estimates. The noise-ceiling serves as a measure of the reliability of the recorded responses, taking into account the inherent noise in the data. To obtain these estimates, we randomly split the trials into two groups, repeating this randomization 100 times. For each experiment, we estimated the noise-ceiling of every site by first computing average normalized net spike rate per image separately for each half of the data, and then measuring the (Spearman–Brown corrected) correlation between the average responses from the two halves. This entire procedure was repeated 100 times for each binary split.

Population Analysis.

Multidimensional scaling. To visualize the similarity relationships between objects in the neuronal population, we employed multidimensional scaling (MDS). The neuronal distances were computed as the average Euclidean distance of the firing rates across neurons for each image pair (alternative distance metrics such as correlation or cosine distance yielded comparable results). We utilized the MDS implementation from the scikit-learn library (implemented in Python). In line with our single-site analysis, we utilized the baseline-subtracted average net spike rates across repetitions of the same image. To ensure equal contribution of each site to the data variance, spike counts were z-standardized.

The MDS procedure was applied to the response matrix of each experiment, with dimensions corresponding to the number of visually responsive sites by the number of individual images. To evaluate the alignment between the two-dimensional embedding and the actual neuronal distances, we calculated the correlation between them. The resulting correlation coefficient is displayed in each multidimensional scaling plot. Notably, the observed high correlation values indicate that the multidimensional scaling plots provide a reasonable approximation of the underlying similarity relationships.

Decoding category information. We used linear decoders to quantify the amount of information about different stimulus categories in the recorded population of neurons on a trial-by-trial basis. We performed this analysis separately for each recording session using every recorded site to prevent bias toward any category. To quantify neural information in a time-resolved manner, firing rates were calculated within 150 ms bins with a sliding window of 10 ms. For each time bin, we trained a logistic regression classifier with L1 regularization and C = 0.1, implemented using the scikit-learn library, to discriminate between categories based on the population firing rate in the same time bin. The specific categories chosen depended on the experiment and hypothesis being tested. To account for class imbalance, we adjusted the logistic regression loss function weights inversely proportional to class frequencies in the input data. For multiclass problems (e.g., Fig. 3*D*), we trained the decoders using a standard "one-vs-rest" approach. Specifically, a linear model was trained to discriminate between each individual category and the remaining categories in a binary fashion. Then, the population response in a specific time bin (the sample) was fed to every individual model and the sample category was attributed to the corresponding category of the model with the higher probability output.

To evaluate the decoding performance, we used the area under the ROC curve metric (AUC), which is insensitive to class imbalance as it is based on the true positive and false positive rates rather than the actual number of instances in each class. Classification AUC was estimated using a 10-fold exemplar-grouped cross-validation approach. In each validation step, we split the trials so that each repetition of a particular image could be in either the training or the test set, with each image exemplar being tested at least once. Before classification, net spike counts were z-standardized to ensure that classification was not influenced by the absolute magnitude of the responses. Training and test sets were not scaled independently but the same scaling was applied to both. To statistically compare decoding AUC to chance performance we computed for each time window the null AUC distribution by shuffling trial indices. For each time window, shuffling was repeated 250 times. Subsequently, actual decoding performance was compared to the null distribution using a one-sided permutation test and by considering only clusters with 5 consecutive significant bins.

Correlation of stimulus representations under shape-preserving image transformations. We wanted to assess the tolerance of the recorded populations to different identity-preserving image transformations in a time-resolved manner. We reasoned that if the neuronal population is tolerant to a specific transformation (e.g., 2D plane rotation) then the representation for each stimulus at the reference condition and the transformed condition should be correlated. To investigate this, we computed the average correlation between each population

vector at a reference orientation and the population vectors corresponding to the same stimulus at a different orientation.

In order to provide reliability of the data given the noise and to provide CI, we binned the data according to each respective transformation and then randomly partitioned the data into two halves. In particular, trials were grouped into S = number of stimuli (variable over experiments), T = 2 pair of transformations (e.g., rotations of 45° and 90°), and H = 2 halves, yielding S × T × H = M total conditions. For each of these conditions, at the evoked response period, we computed the average population vector $r_{S,T,H}$. Finally, the average correlation of the population response across transformations was computed as follows:

$$\rho_{cross} = \frac{1}{STH} \sum_{i=1}^{H} \sum_{j=1}^{H} \sum_{s=1}^{S} corr\left(r_{s,1,i} - \langle r_{s,1,i}\rangle_S, r_{s,2,j} - \langle r_{s,2,j}\rangle_S\right),$$

where $\langle * \rangle_S$ corresponds to the average response vector across the set of stimuli S. Namely, for each set of S population vectors corresponding to a particular half of the data H and transformation T, we subtracted the mean across transformations to center the population vectors around zero. Therefore, ρ_{cross} measures to what extent stimulus population representations are similarly organized around their mean across the two transformations.

We computed the average correlation of each population vector with itself across the two halves of the trials with the aim of obtaining an upper bound to ρ_{cross} given the degree of noise in the data:

$$\rho_{self} = \frac{1}{ST} \sum_{s=1}^{S} \sum_{t=1}^{T} corr\left(r_{s,t,1} - \langle r_{s,t,i}\rangle_S, r_{s,t,2} - \langle r_{s,t,2}\rangle_S\right).$$

This analysis included only sites that responded significantly to at least one of the conditions (or rotations).

To obtain CI for ρ_{cross} and ρ_{self}, we randomly split the trials for each condition (stimulus–transformation) into two groups, repeating this randomization 1,000 times. The average firing rates across these sampled trials provided the mean population vector for that condition on that iteration.

To determine the significance of the computed ρ_{cross} and ρ_{self} values, we conducted permutation tests. In these tests, we shuffled the labels of the trials associated with the analyzed transformations (reference and target) to create a null distribution. For each iteration of the permutation test, we followed the same resampling procedure described earlier to obtain an average metric value for that iteration. Specifically, we performed 1,000 permutation iterations, with 1,000 resampling iterations per permutation iteration.

Decoding Generalization of Category Information Under Certain Image Transformations. We conducted a generalization decoding analysis to investigate the ability of the recorded neuronal populations to encode category information despite certain image transformations. We reasoned that if the population is tolerant to certain image transformations, responses to intact and transformed images would share a common population representation. In such case, a decoder trained on intact (original) images should exhibit significant classification performance when tested with transformed stimuli.

To evaluate the decoding performance for silhouette images, we performed a cross-validation training procedure where we trained the decoders using a subset of the intact images and we tested on held-out images for both intact images and silhouette versions; the held-out images belonged to the same object exemplars for both intact and silhouette images. In the case of decoding generalization in the abstract body test (*SI Appendix*, Fig. S8), the images from the intact bodies and the abstract body images came from entirely different distributions. To assess significance for the decoding performance of the transformed sets, we performed permutation tests where the null distribution was created by shuffling the labels of both the training set and the generalization test set.

V1 Model Control. To investigate whether the observed decoding results could arise from neural activity tuned for low-level visual features, we employed a V1 model [VOneBlock (10)] consisting of Gabor filters. In brief, the VOneBlock model is based on the classical linear–nonlinear-Poisson (LNP) model, consisting of a

Gabor filter bank (GFB), simple and complex cell nonlinearities, and Poisson-like stochasticity.

To simulate a V1-like dataset for each experiment, we passed the stimuli through the V1 model. To account for the variability in the recorded neural responses, we simulated the repetitions of single trials by passing the same image multiple times, equal to the number of repetitions per stimulus in the original experiment. As the simulated data were subject to Poisson-like noise, we obtained different activation patterns for the same image, resembling the statistical characteristics of real neural data. The model convolved the input images with 512 Gabor filters, each having different orientations, sizes, shapes, and spatial frequencies. This process generated a 56×56 spatial map for each image, resulting in a stack of 512 output images, one for each type of Gabor filter.

To ensure a fair comparison between the neural data and the model results, we performed a subsampling procedure. We repeated this process 1,000 times, randomly selecting a number of V1 units equal to the visually responsive sites specific to the array and experiment. For each iteration, we assessed decoding performance using group cross-validated linear decoding, following the same procedure as in the neural data analysis. Finally, we assessed the significance of the comparison between neural data and the V1 model results by calculating P-values as follows:

$$p = \frac{C + 1}{N_{permutations} + 1}.$$

where C is the number of permutations whose performance was higher than the one obtained with neural data.

Data, Materials, and Software Availability. Data Files data have been deposited in Dryad (https://doi.org/10.5061/dryad.2v6wwpzz0) (50).

Author affiliations: ªResearch group Experimental Neurosurgery and Neuroanatomy, Katholieke Universiteit Leuven, and the Leuven Brain Institute, Leuven B-3000, Belgium; ᵇLaboratory for Neuro- and Psychophysiology, Department of Neurosciences, Katholieke Universiteit Leuven and the Leuven Brain Institute, Leuven B-3000, Belgium; ᶜDepartment of Neurosurgery, Universitaire Ziekenhuizen Leuven, Katholieke Universiteit Leuven, Leuven B-3000, Belgium; ᵈDepartment of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139; ᵉMcGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA 02139; ᶠThe Center for Brains, Minds and Machines, Massachusetts Institute of Technology, Cambridge, MA 02139; ᵍLaboratory for Epilepsy Research, Katholieke Universiteit Leuven, Leuven B-3000, Belgium; ʰDepartment of Psychology and Cognitive Science, University of Trento, Trento 38068, Italy; and ⁱLaboratory for Biological Psychology, Katholieke Universiteit Leuven, Leuven B-3000, Belgium

1. M. V. Peelen, P. E. Downing, The neural basis of visual body perception. *Nat. Rev. Neurosci.* **8**, 636–648 (2007).
2. R. Vogels, More than the face: Representations of bodies in the inferior temporal cortex. *Ann. Rev. Vis. Sci.* **8**, 383–405 (2022).
3. G. Pourtois, M. V. Peelen, L. Spinelli, M. Seeck, P. Vuilleumier, Direct intracranial recording of body-selective responses in human extrastriate visual cortex. *Neuropsychologia* **45**, 2621–2625 (2007).
4. P. E. Downing, Y. Jiang, M. Shuman, N. Kanwisher, A cortical area selective for visual processing of the human body. *Science* **293**, 2470–2473 (2001).
5. J. B. Julian, E. Fedorenko, J. Webster, N. Kanwisher, An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *Neuroimage* **60**, 2357–2364 (2012).
6. I. D. Popivanov, J. Jastorff, W. Vanduffel, R. Vogels, Stimulus representations in body-selective regions of the macaque cortex assessed with event-related fMRI. *Neuroimage* **63**, 723–741 (2012).
7. D. R. Kenshalo, B. M. David Green, J. A. Swets John, Signal detection theory and psychophysics. *J. Acoust. Soc. Am.* **42**, 578–578 (1967).
8. I. D. Popivanov, J. Jastorff, W. Vanduffel, R. Vogels, Heterogeneous single-unit selectivity in an fMRI-defined body-selective patch. *J. Neurosci.* **34**, 95–111 (2014).
9. M. V. Peelen, P. E. Downing, The neural basis of visual body perception. *Nat. Rev. Neurosci.* **8**, 636–648 (2007).
10. J. Dapello *et al.*, "Simulating a primary visual cortex at the front of CNNs improves robustness to image perturbations" in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, H. Lin, Eds. (Curran Associates, Inc., 2020), vol. 33, pp. 13073–13087.
11. P. E. Downing, M. V. Peelen, Body selectivity in occipitotemporal cortex: Causal evidence. *Neuropsychologia* **83**, 138–148 (2016).
12. J. C. Taylor, A. J. Wiggett, P. E. Downing, Functional MRI analysis of body and body part representations in the extrastriate and fusiform body areas. *J. Neurophysiol.* **98**, 1626–1633 (2007).
13. C. Urgesi, B. Calvo-Merino, P. Haggard, S. M. Aglioti, Transcranial magnetic stimulation reveals two cortical pathways for visual body processing. *J. Neurosci.* **27**, 8023–8030 (2007).
14. J. Schrouff *et al.*, Fast temporal dynamics and causal relevance of face processing in the human temporal cortex. *Nat. Commun.* **11**, 656 (2020).
15. S. Bracci, M. Ietswaart, M. V. Peelen, C. Cavina-Pratesi, Dissociable neural responses to hands and non-hand body parts in human left extrastriate visual cortex. *J. Neurophysiol.* **103**, 3389–3397 (2010).
16. S. Bracci, A. Caramazza, M. V. Peelen, Representational similarity of body parts in human occipitotemporal cortex. *J. Neurosci.* **35**, 12977–12985 (2015).
17. T. Orlov, T. R. Makin, E. Zohary, Topographic representation of the human body in the occipitotemporal cortex. *Neuron* **68**, 586–600 (2010).
18. K. Tanaka, Inferotemporal cortex and object vision. *Ann. Rev. Neurosci.* **19**, 109–139 (1996).
19. Z. Kourtzi, N. Kanwisher, Representation of perceived object shape by the human lateral occipital complex. *Science* **293**, 1506–1509 (2001).
20. N. Caspari *et al.*, Fine-grained stimulus representations in body selective areas of human occipito-temporal cortex. *Neuroimage* **102**, 484–497 (2014).
21. R. Vogels, More than the face: Representations of bodies in the inferior temporal cortex. *Ann. Rev. Vis. Sci.* **8**, 383–405 (2022).
22. I. D. Popivanov, J. Jastorff, W. Vanduffel, R. Vogels, Tolerance of macaque middle STS body patch neurons to shape-preserving image transformations. *J. Cogn. Neurosci.* **27**, 1001–1016 (2015).
23. S. Kumar, I. D. Popivanov, R. Vogels, Transformation of visual representations across ventral stream body-selective patches. *Cereb. Cortex* **29**, 215–229 (2019).
24. J. J. DiCarlo, J. H. R. Maunsell, Using neuronal latency to determine sensory–motor processing pathways in reaction time tasks. *J. Neurophysiol.* **93**, 2974–2986 (2005).
25. P. L. Croxson, S. J. Forkel, L. Cerliani, M. Thiebaut de Schotten, Structural variability across the primate brain: A cross-species comparison. *Cereb. Cortex* **28**, 3829–3841 (2018).
26. P. Bao, L. She, M. McGill, D. Y. Tsao, A map of object space in primate inferotemporal cortex. *Nature* **583**, 103–108 (2020).
27. A. H. Bell *et al.*, Relationship between functional magnetic resonance imaging-identified regions and neuronal category selectivity. *J. Neurosci.* **31**, 12229–12240 (2011).
28. P. E. Downing, M. V. Peelen, The role of occipitotemporal body-selective regions in person perception. *Cogn. Neurosci.* **2**, 186–203 (2011).
29. P. Bashivan, K. Kar, J. J. DiCarlo, Neural population control via deep image synthesis. *Science* **364**, eaav9436 (2019).
30. C. R. Ponce *et al.*, Evolving images for visual neurons using a deep generative network reveals coding principles and neuronal preferences. *Cell* **177**, 999–1009.e10 (2019).
31. S. A. Cadena *et al.*, Deep convolutional models improve predictions of macaque V1 responses to natural images author summary. *PLoS Comput. Biol.* **15**, e1006897 (2017). 10.12751/g-node.2e31e3.
32. C. Zhuang *et al.*, Unsupervised neural network models of the ventral visual stream. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2014196118 (2021).
33. S. M. Khaligh-Razavi, N. Kriegeskorte, Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS Comput. Biol.* **10**, e1003915 (2014).
34. U. Güçlü, M. A. J. van Gerven, Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *J. Neurosci.* **35**, 10005–10014 (2015).
35. N. A. Ratan Murty, P. Bashivan, A. Abate, J. J. DiCarlo, N. Kanwisher, Computational models of category-selective brain regions enable high-throughput tests of selectivity. *Nat. Commun.* **12**, 1–14 (2021).
36. B. Long, C. P. Yu, T. Konkle, Mid-level visual features underlie the high-level categorical organization of the ventral stream. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E9015–E9024 (2018).
37. R. Geirhos *et al.*, "Generalisation in humans and deep neural networks" in *Advances in Neural Information Processing Systems*, S. Bengio *et al.*, Eds. (Curran Associates Inc, 2018), **vol. 31**.
38. T. Serre, Deep learning: The good, the bad, and the ugly. *Ann. Rev. Vis. Sci.* **5**, 399–426 (2019).
39. K. Kar, J. Kubilius, K. Schmidt, E. B. Issa, J. J. DiCarlo, Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. *Nat. Neurosci.* **22**, 974–983 (2019).
40. T. Decramer *et al.*, Single-cell selectivity and functional architecture of human lateral occipital complex. *PLoS Biol.* **17**, e3000280 (2019).
41. T. Decramer *et al.*, Single-unit recordings reveal the selectivity of a human face area. *J. Neurosci.* **41**, 9340–9349 (2021).
42. L.-E. Martinet *et al.*, Human seizures couple across spatial scales through travelling wave dynamics. *Nat. Commun.* **8**, 14896 (2017).
43. W. Truccolo *et al.*, Single-neuron dynamics in human focal epilepsy. *Nat. Neurosci.* **14**, 635–641 (2011).
44. E. H. Smith *et al.*, The ictal wavefront is the spatiotemporal source of discharges during spontaneous human seizures. *Nat. Commun.* **7**, 11098 (2016).
45. V. Bougou *et al.*, Neuronal tuning and population representations of shape and category in human visual cortex. *Nat. Commun.* **15**, 4608 (2024).
46. K. Matić, H. Op de Beeck, S. Bracci, It's not all about looks: The role of object shape in parietal representations of manual tools. *Cortex* **133**, 358–370 (2020).
47. V. Willenbockel *et al.*, Controlling low-level image properties: The SHINE toolbox. *Behav. Res. Methods* **42**, 671–684 (2010).
48. V. Fonov *et al.*, Unbiased average age-appropriate atlases for pediatric studies. *Neuroimage* **54**, 313–327 (2011).
49. A. O. Blenkmann *et al.*, Ielectrodes: A comprehensive open-source toolbox for depth and subdural grid electrode localization. *Front. Neuroinform.* **11**, 1–16 (2017).
50. J. G. Ramirez, P. Janssen, T. Theys, Data from Intracortical recordings reveal the neuronal selectivity for bodies and body parts in the human visual cortex. Dryad. https://datadryad.org/stash/dataset/doi:10.5061/dryad.2v6wwpzz0. Deposited 13 October 2024.