

Supplementary Information for

Mid-level visual features underlie the high-level categorical organization of the ventral stream

Bria Long, Chen-Ping Yu, & Talia Konkle

Bria Long

Email: brialorelle@gmail.com

This PDF file includes:

1. Extended Methods
 - a. fMRI preprocessing
 - b. MRI acquisition
 - c. fMRI experiment design details
 - d. Retinotopy protocol
 - e. Preference maps correlation details
 - f. Posterior-to-anterior correlation details
 - g. Predictive modeling: feature extraction

2. Extended Materials: Texform stimuli details
 - a. Stimulus set construction: **Figure S1**
 - b. Texform Selection Details
 - c. Basic-level norming task example; texform classifiability vs. basic-level recognizability. **Figure S2A/B**
 - d. Animacy/size classification task and classifiability groups: **Figure S3A/B**
 - e. Curvature rating task and results: **Figure S4**
 - f. Post-scan recognizability of the texforms: **Figure S5**

3. Supplement to Experiment 1
 - a. Voxel mask construction: **Figure S6**
 - b. Group-level topographies: **Figure S7**
 - c. Single-subject topographies: **Figure S8**

- d. Posterior-to-anterior preference scatterplots, **Figure S9**
 - e. Overall response differences between originals and texforms: **Figure S10**
 - f. Comparison to Konkle & Caramazza, 2013: **Figure S11**
4. Supplement to Experiment 2
- a. Eye-tracking stability: **Figure S12**
 - b. Voxel mask construction: **Figure S13**
 - c. Group-level conjunction topographies: **Figure S14**
 - d. Single-subject conjunction topographies: **Figure S15**
5. Supplement to Predictive Modeling
- a. Voxel mask construction: **Figure S16**
 - b. Modeling results in anterior to posterior regions: **Figure S17, Figure S18**
 - c. Early visual cortex results and corresponding **Figure S19**

Other supplementary materials for this manuscript include the following:

All stimuli, pre-processed data, and main analysis code for this paper are available at the Open Science Repository for this project, <https://osf.io/69pbd/>, which is also linked to a GitHub codebase for generating texforms. Raw fMRI data is available on request.

1. Extended Methods

fMRI Data Preprocessing. Functional data were analyzed using Brain Voyager QX software and MATLAB. Preprocessing included slice scan-time correction, 3D motion correction, linear trend removal, temporal high-pass filtering (0.01 Hz cutoff), spatial smoothing (4 mm FWHM kernel), and transformation into Talairach (TAL) coordinates. Two subjects had one run in which they moved more than 0.5 mm within 2 seconds (1 TR) and these runs were discarded from analysis. The cortical surface of each subject from the high-resolution T1-weighted anatomical scan, acquired with a 3D MPRAGE protocol. To do so, we used the default segmentation procedures in FreeSurfer. Surfaces were then imported into Brain Voyager and inflated using BV surface module. Gray matter masks were defined in the volume and were constructed based on the FreeSurfer segmentations.

General linear models (GLMs) were computed at the single subject level for texforms and original runs separately, both for the four main conditions (big animals, big objects, small animals, and small objects) as well as separately for the full set of nested conditions (each category x each classifiability level, 24 conditions total). GLMs included square-wave regressors for each condition's presentation times, convolved with a gamma function to approximate the hemodynamic response, fit to voxel-wise time course data with percent signal change normalization and correction for serial correlations. In Experiment 2, GLMs were fit eight main conditions of interest: each combination of category (big animals, big objects, small animals, and small objects) and visual field presentation (upper, lower) separately for texforms and originals.

MRI acquisition. Imaging data were collected using a 32-channel phased-array head coil with a 3T Siemens Prisma fMRI Scanner at the Harvard Center for Brain Sciences. High-resolution T1-weighted anatomical scans were acquired using a 3D MPRAGE protocol (176 sagittal slices; FoV = 256 mm; 1x1x1 mm voxel resolution; gap thickness = 0 mm; TR = 2530 ms; TE = 1.69 ms; flip angle = 7 degrees). For functional runs, blood oxygenation level-dependent (BOLD) contrast was obtained using a gradient echo-planar T2* sequence (84 oblique axial slices acquired at a 25° angle off of the anterior commissure-posterior commissure line; FoV = 204 mm; 1.5x1.5x1.5 mm voxel resolution; gap thickness = 0 mm, TR = 2000 ms; TE = 30 ms; flip angle = 80 degrees; multi-band acceleration factor = 3).

fMRI Experiment Design. Each run had twelve 6s blocks for each condition (big animals, big objects, small animals, small objects), with 10s rest periods interleaved every four blocks. Each block consisted of six images (5 unique images and 1 repeat) each presented for 800ms followed by a 200ms blank. Further, each block contained images from one of the six classifiability levels. Each classifiability level for each condition was shown twice per run. Thus, each texform image was shown twice during a run and 8 times over the entire experiment. All images were presented in isolation on a uniform gray background. This design choice allowed us to analyze neural responses in both a high-powered, four-condition design as well as a moderately powered 24-condition design for the predictive modeling analysis. Note, however, this does not allow us to model responses to individual texforms or their corresponding recognizable images.

In Experiment 1, each image subtended 10.36° x 10.36° visual angle centered at fixation. In Experiment 2, each block of images could appear either above or below fixation (6.92° x 6.92° degrees of visual angle, bottom edge .86° degrees from center). These positions were counterbalanced across blocks such that, for each level of classifiability and condition, one block was presented in the upper visual field and the other block was presented in the lower visual field. Participants were instructed that maintaining fixation was more important than task performance, and fixation was monitored online using an EyeLink

1000 eye-tracker. Participants were calibrated to the eye-tracker at the beginning of the experiment and were recalibrated every 2-3 runs as needed. See **Figure S12** for fixation heatmaps for each participant.

Retinotopy Protocol. Additionally, participants completed a retinotopy protocol in order to define early visual areas V1-V3. Observers viewed bands of flickering checkerboards in a blocked design. The conditions included vertical meridian bands ($\sim 22^\circ \times 1.7^\circ$), horizontal meridian bands ($\sim 22^\circ \times 1.7^\circ$), iso-eccentricity bands covered by a central ring (radius $\sim 1.2^\circ$ to 2.4°), a peripheral ring (radius $\sim 5.7^\circ$ to 9.3°), and an extra wide peripheral ring (inner radius $\sim 9.3^\circ$, filling the extent of the screen). In Experiment 2, the vertical and horizontal meridian bands were replaced with wedges. The apex of each wedge was at fixation and the base extended to $\sim 22^\circ$ in the periphery, and the checkerboard patterns flickered at 6 Hz. Each block was 6 seconds, within which the checkerboard cycled at 8 Hz between states of black-and-white, randomly colored, white-and-black, and random colored. In each 4.4-min run (142 volumes), the 5 visual field band conditions and 1 fixation condition were repeated 7 times with their order randomly permuted within each repetition. Each run started and ended with a 6 s fixation period. Participants' task was to maintain fixation and press a button every time the fixation dot turned red, which happened once per block. Using data from this retinotopy protocol, early visual regions (V1-V3) were defined by hand on inflated brain guided by the contrast of horizontal vs. vertical meridians (see (1)).

Preference Map Correlation Details. Two comparisons were used to assess map-correlation robustness. First, we compared map correlations to a shuffled voxel baseline. For each subject, the spatial position of texform voxels was shuffled and then correlated with the unshuffled original preference map. This was repeated 1000 times, yielding a chance distribution for each subject, from which a p-value was computed based on how often the simulated shuffled values were greater than the observed map correlation. Second, we considered the map correlations relative to an estimated noise ceiling. To do so, in each subject, texform preference maps were correlated between odd and even runs, yielding a texform map split-half correlation. The same analysis was repeated for originals. If any of these odd-even correlations was less than zero (i.e., a negative correlation), we substituted this value with zero; this occurred in one subject for the object size comparison. Given these split-half texform and original map correlations were estimated with half the power of the texform-original map comparison, we used the Spearman Brown prophecy formula to approximately adjust the reliabilities ($N \cdot \text{observed reliability} / 1 + (N-1) \cdot \text{observed reliability}$, where $N = 2$ as we divided the data in half). Then, the noise ceiling for the texform-original map correlation was computed separately for each participant, as the square root of the product of these corrected reliabilities.

Posterior-to-Anterior Correlations. To assess whether there was a difference in the overall strength of the original animacy preferences vs. the texform animacy preferences along the posterior-to-anterior gradient, we computed difference scores for each anatomical section for each participant. We then performed a simple rank correlation between ascending anatomical sections (i.e., 1,2,3,4,5) and these difference scores (originals – texforms) in each subject. A rank correlation metric was used to assess whether originals generate greater animacy preferences along this posterior to anterior gradient, without assuming a meaningful relationship with the TAL-Y coordinates of the anatomical sections. Finally, we asked whether these rank correlations were above zero at the group level by performing a one-sample t-test over subjects. The same procedure was repeated for the object size distinction. For visualization purposes, in **Figure S9A-C**, we also defined group-level anatomical sections based on the Group GLM activations, with the scatter plots showing voxel response preferences for animacy and size dimensions based on the group GLM beta fits.

Predictive Modeling Feature Spaces.

Gabor & Gist Models. Gabor features were extracted in an 8 x 8 grid over the original, recognizable images (440 x 440 pixels) at three different 3 scales, with 8, 6, and 4 oriented Gist per scale, respectively

(Oliva & Torralba, 2006). GIST model features were extracted by taking the first 20 principle components of this Gabor feature matrix. In both cases, these features were then averaged across the five images in each nested classifiability group presented during the fMRI experiment. The squared Euclidean distance along each feature was used to construct feature RDMs for use in the predictive modeling procedure, and all dissimilarities were scaled between 0 and 1, yielding a 276 x 896 feature vector for Gabor features and a 276 x 20 feature vector for Gist features.

Texture Synthesis Model (Freeman & Simoncelli, 2011): The texture synthesis model has 10 feature classes (corresponding to pixel statistics, weighted marginal statistics, simple cell responses, complex cell responses, cross-position correlations (i.e., autocorrelation) within scales computed separately for simple and complex cells, cross-orientation correlations computed separately for simple and complex cells, and cross-scale correlations computed separately for simple and complex cells). Features were included if they (1) had any variance across the images ($SD > 0$) and (2) were calculated within pooling windows tiling the depicted item (see **Figure S1** for an illustration of the pooling windows). The values for each feature were then z-scored across the 120 images, and then averaged over the five images in each classifiability group. Each feature was converted to an RDM using squared Euclidean distance, and all dissimilarities were scaled between 0 and 1, generating a 276 x 20,914 feature matrix.

Behavioral Ratings–Animacy/Size: For texforms, feature RDMs were constructed based on the behavioral animacy and size classifiability scores. Note these are the same scores used to group the texforms into the nested design. These experiments yielded a vector corresponding to participants ability to classify each texform as an animal (range: 0-1, where 1 = always classified as an animal, and 0 = never classified as an animal) and their ability to classify each texform according to their size in the real world (range: 0-1, where 1 = always classified as big in the real-world, and 0 = never classified as big in the real world). These scores were averaged according to the 24 nested conditions presented during the experiment, yielding a 24 x 1 vector for animacy and a 24 x 1 vector for size for texforms and for originals. We then took the squared Euclidean distance of each 1-dimensional feature vector and all dissimilarities were scaled between 0 and 1; the final feature matrix was a 276 x 2 feature matrix.

For original images, feature RDMs were constructed using their actual animacy/size in real-world. These yielded a vector corresponding to the actual animacy of the recognizable image (1 = animate, and 0 = inanimate), and a vector corresponding the actual size of the object in the real world (1 = big in the real-world, and 0 = small in the real world). We then took the squared Euclidean distance of each 1-dimensional feature vector and all dissimilarities were scaled between 0 and 1; the final feature matrix was a 276 x 2 feature matrix.

Behavioral Ratings–Curvature: Behavioral ratings on Amazon Mechanical Turk were obtained to assess the perceived curvature of both the texforms and the originals; 30 participants rated the curvature of the 120 texforms, and another 30 participants rated the curvature of their 120 corresponding original images. Participants were asked, “How boxy or curvy is the thing depicted in this image?” and asked to respond using a 1-5 scale (1: Very curvy, 2: Mostly curvy, 3: Equally boxy and curvy, 4: Mostly boxy, 5: Very boxy). See **Figure S4A** for an illustration of the task. These ratings were averaged across participants, and then averaged across the five images in each classification group. This yielded two 24 x 1 vectors corresponding to the average perceived curvature of each group of texforms and of each group of original images. The squared Euclidean distance of all pairwise comparisons of these conditions was computed separately for texforms and originals, yielding two 276 x 1 feature vectors for curvature for modeling responses to texforms and originals; all dissimilarities were scaled between 0 and 1. See **Figure S4B** for a visualization of this data and a comparison of the curvature ratings between texforms and recognizable images.

CNN Features (Texforms & Originals): The AlexNet architecture (2) as was trained using the conventional image classification task using the ImageNet dataset. The standard AlexNet training regime was adopted using a public code package (<https://github.com/soumith/imagenet-multiGPU.torch>) that was optimized for multi-threaded CNN training in Torch7. Specifically, stochastic gradient descent (SGD) optimization was used with 0.9 momentum, an initial learning rate of 0.02, and weight decay of 0.0005. Both the learning rate and the weight decay follow a pre-defined decreasing schedule (see train.lua from the code package) using a mini-batch size of 128, with 10,000 mini-batches per epoch over a total of 55 training epochs. Standard data augmentation such as random horizontal flips and random 224-by-224 crops were performed during training.

Using the fully-trained network, CNN features were extracted from each unit in the CNN from both original and texform image sets. Specifically, for each image and each convolutional filter, we computed the summed activation map of the filter (an m -by- m map where m is the output size of the convolutional layer), accounting for border effects by setting to zero all values in the activation map within 10% of the four edges. This procedure yielded five feature matrices for the original images of 120-by-64, 120-by-192, 120-by-384, 120-by-256, and 120-by-256, corresponding to each of the five convolutional layers, and another set corresponding to the texform images. For the two fully-connected layers (layer 6 & 7), the activation level to each image was direct computed (no global summation required), resulting in two feature matrices of 120-by-4096, corresponding to layer 6 and 7, and another set for texform images.

Each feature matrix was normalized by dividing the rows with its L2-norm, and the rows were averaged over the five images in each classifiability group. Finally, for each feature (column) of each feature matrix, we computed the pairwise squared Euclidean distance between all of the 24 conditions, yielding five 276-by- m representational dissimilarity matrices, where m is the number of convolutional filters for the corresponding layer. All dissimilarities were then scaled between 0 and 1. These RDMs were used to perform feature modeling of the individual CNN layers.

2. Extended Materials: Texform stimuli details

Here, we provide additional details on (a) how the stimuli set was constructed (**Fig. S1**) (b) an overview of how basic-level recognition was assessed (**Fig. S2A**), (c) how classifiable the texforms are by their animacy/size, how these were used to form the nested groupings, and relationship between basic-level recognition and classifiability (**Fig. S3, Fig. S2B**), (d) the perceived curvature of the texforms and recognizable images (**Fig. S4**) and (e) how recognizable the texforms were after the neuroimaging session (**Fig. S5**).

Texform generation overview

For every image in super set (240 images):

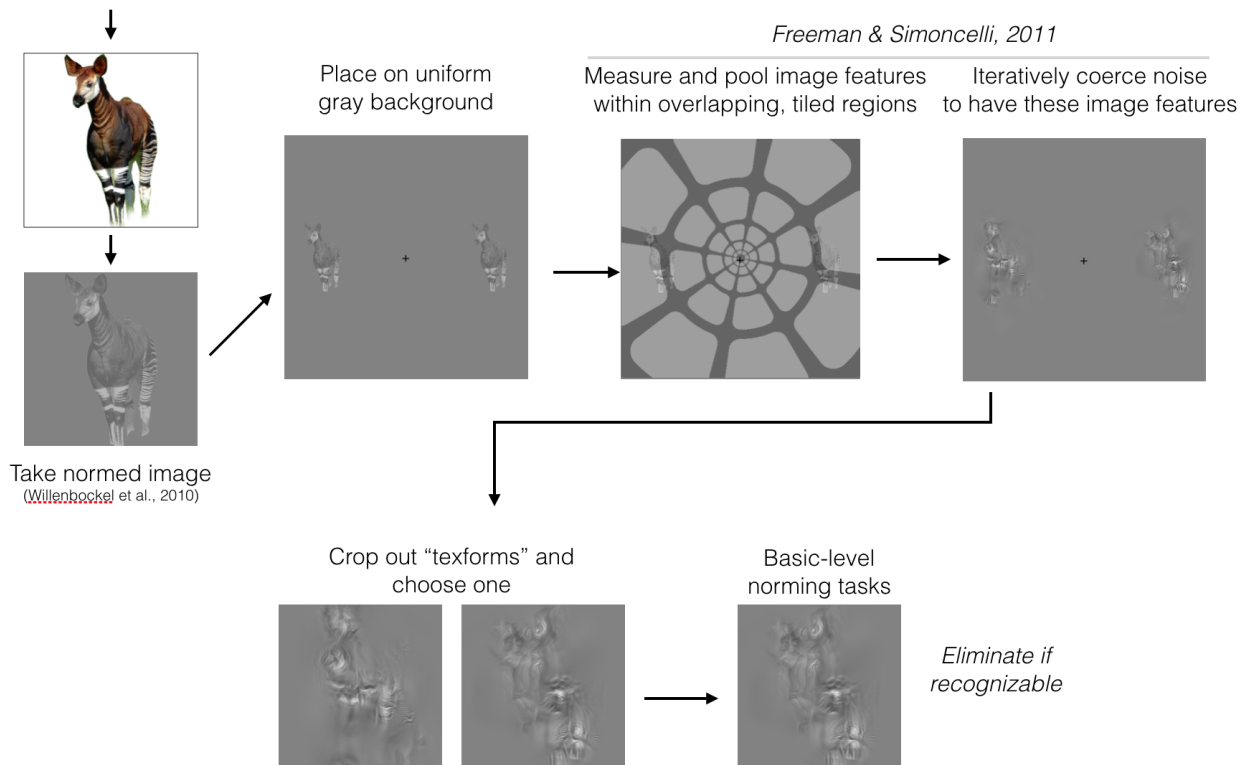



Fig. S1. Overview of the procedure used to generate texforms. Original, color, images were converted to grayscale and matched for overall luminance and contrast within the superset of 240 images. Next, these normed images were placed in the “periphery” of the model on the uniform gray background. First and second-order image statistics (3) were measured and pooled within overlapping, tiled regions illustrated here (pooling window parameters, scaling = .5, AR = 1). Next, synthetic stimuli were generated by coersing random noise to have the similar image statistics within these pooling windows. The procedure was run for 50 iterations using a variant of gradient descent. This produced a synthesized image with two texforms, which were then cropped out. Norming tasks were then used to select a set of texforms that were unrecognizable at the basic-level (see below).

Texform Selection Details. Online recognition experiments were run to assess how recognizable each texform was. First, 18 participants guessed the identity of each of 240 texform images. Next, six new participants assessed the validity of these guesses. These participants were presented with the original images and all of the texform guesses and judged whether each guess could be “used to correctly describe” the original image (see **Figure S2A**). The proportion of guesses accepted as correct yielded a basic-level identification score for each image. Images in which a rater accepted more than 3/18 responses as correct were removed. Next, 120 texforms and their corresponding originals were selected (30 images per category), with the constraints that the categories did not significantly differ in either aspect ratio or pixel area; all $p \geq .1$). On average, these 120 texforms were identified at the basic level <3% of the time. Finally, the overall luminance and contrast levels across all 240 images (120 texforms, 120 originals) equated using the SHINE toolbox (4) and the edges of all of images were blurred so that they gradually faded into their backgrounds.

A. Basic-level identification task

18 raters were asked to describe a scrambled version of this image, and their responses are below.

Please select any and ALL of their responses that could be used to correctly describe this image. If none of the responses are appropriate, please leave all checkboxes blank.



<input type="checkbox"/> 'dolphin'	<input type="checkbox"/> 'shell'
<input type="checkbox"/> 'skunk'	<input type="checkbox"/> 'pathways'
<input type="checkbox"/> 'flower'	<input type="checkbox"/> 'fig'
<input type="checkbox"/> 'child'	<input type="checkbox"/> 'person bending over'
<input type="checkbox"/> 'Elephant'	<input type="checkbox"/> 'chair'
<input type="checkbox"/> 'rose'	<input type="checkbox"/> 'chefs hat'
<input type="checkbox"/> 'jellyfish'	<input type="checkbox"/> 'ear'
<input type="checkbox"/> 'A picture of a bicycle'	<input type="checkbox"/> 'headphones'
<input type="checkbox"/> 'pear'	<input type="checkbox"/> 'a face'

B. Basic-level identification

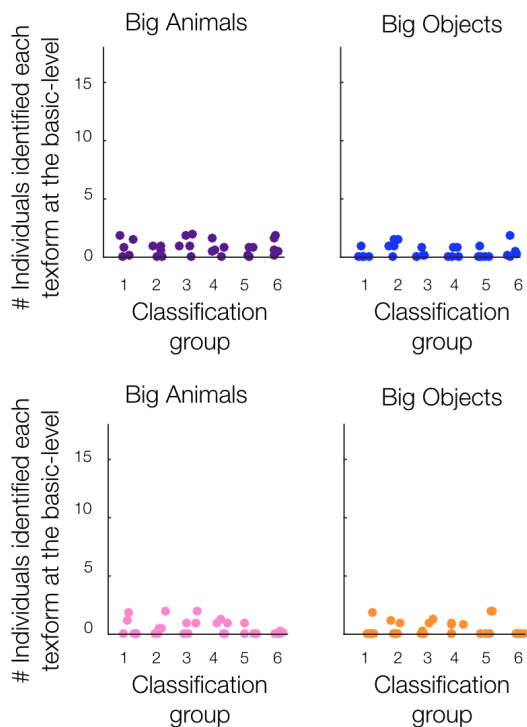
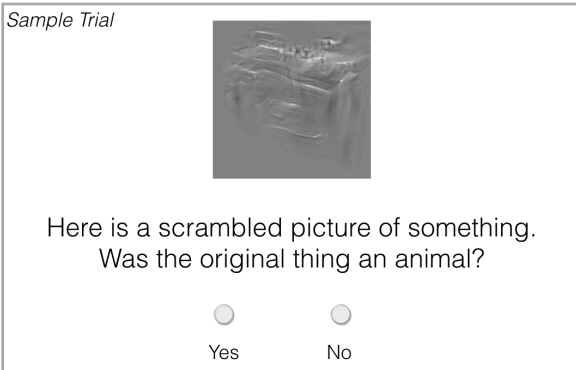


Figure S2. (A). Example trial from the basic-level identification task; raters determined whether the guesses could be used to describe the original image from which the texform was generated. (B). Basic-level identification rate from the 18 norming participants as a function of animacy/size classification group for each object category; each point represents a texform image. X-axis position is jittered to show all points.

A. Animacy/Size Classification Tasks

Separate participants for each question (N=16 each)



Questions

Animacy: Was the original thing an animal?

Animacy: Was the original thing a man-made object?

Size: Was the original thing big enough to support a human?

Size: Was the original thing small enough to hold with one or two hands?

B. Animacy/Size Classification Results

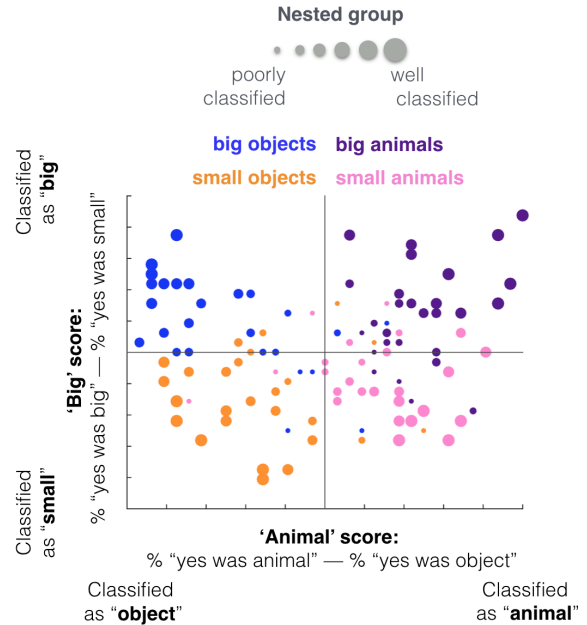
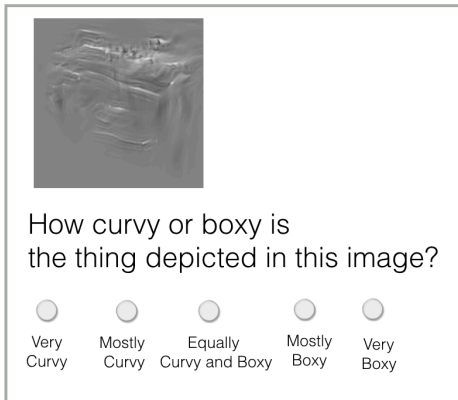


Figure S3. (A) Schematic of the animacy/size classification tasks. (B) The classifiability of each image is plotted for animacy (x-axis) and real-world size (y-axis); each dot corresponds to a texform image. The position of the dot reflects its classifiability score on both axes, the color of the dot indicates the actual condition of the texform (big/small animal/object), and the size of the dot indicates which of the 6 classifiability groups it was assigned to, where larger dots represent groups of texform images that were better classified by their animacy and real-world size.

A. Curvature Ratings: Task

For both texforms and originals
Separate participants (N=30 each)



B. Curvature Ratings: Results

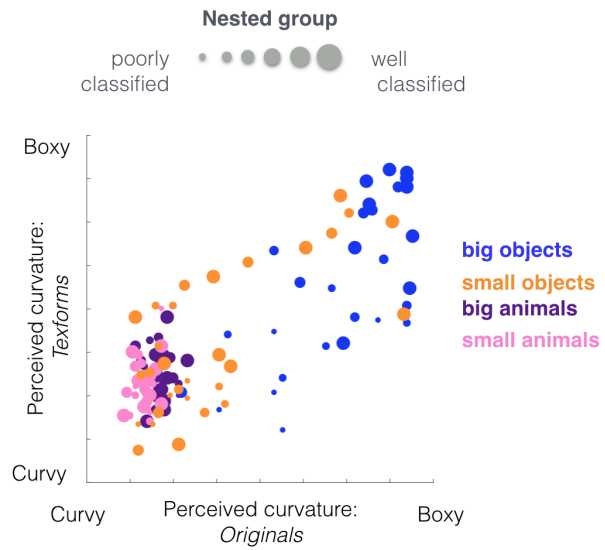
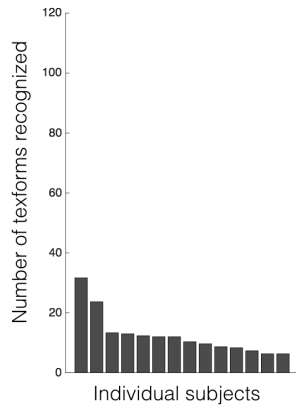


Figure S4. (A) Schematic of the perceived curvature task. (B). The perceived curvature of the texforms (y-axis) is plotted as a function of the perceived curvature of the recognizable, original images (x-axis); larger dots represent groups of images that were better classified by their animacy and real-world size.

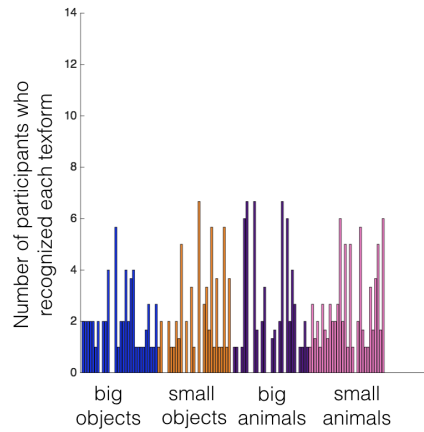
Post-test basic-level recognition

After seeing texforms 8x then originals 8x

A. By participant



B. By item



C. Recognition by classifiability

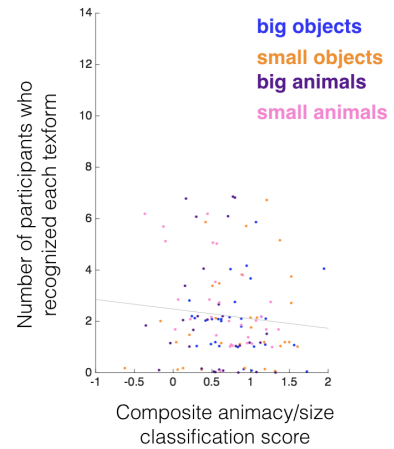


Figure S5. *Post-test* recognition results for the neuroimaging participants. After the scanning session, participants were told that the texforms actually were generated from real-world objects, and then completed a task in which they guessed what each texform might be. Note that all observers had seen each texform image eight times, then each original image 8 times, while in the scanner, before this test was taken. Three naïve observers rated whether the participants' texform guesses could be used to describe the original images, and they were told to be generous with what they counted as correct. (A) The number of texforms recognized by each participant is plotted. (B) The number of participants who recognized each texform is plotted; texforms are ordered according to their condition and composite classifiability score. (C) The item effects in B are plotted by the classifiability score of each item. We did not find strong evidence that the more classifiable texforms were also the ones that were more likely to be recognized after the scanning session; if anything, the trend was in the opposite direction.

3. Supplement to Experiment 1

Here, we first (a) provide additional details on how active OTC voxels were selected (**Fig. S6**), (b) show group preference maps (**Fig. S7**) and all individual subject preference maps (**Fig. S8**) for the animacy and object size distinctions, (c) plot group tripartite maps for direct comparison with Konkle & Caramazza (2013) (5) (**Fig. S9**), (d) plot posterior-to-anterior scatterplots of group-level animacy and object size preferences (**Fig. S10**), and (e) illustrate overall differences in response magnitude to texforms vs. recognizable images (**Fig. S11**).

Defining active OTC

Single subject example - defined individually for each subject

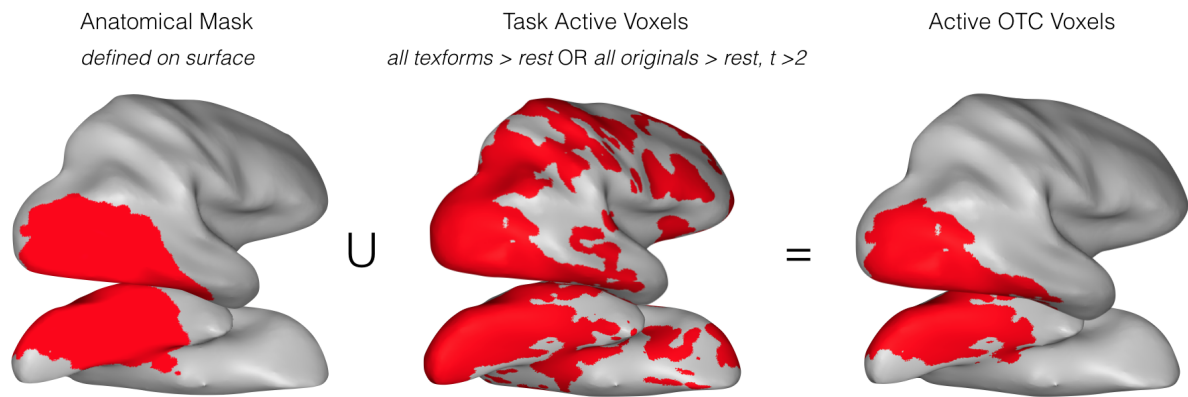
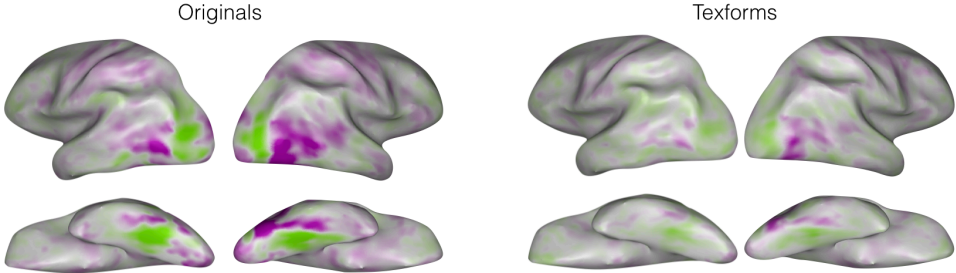


Figure S6. Schematic of how active OTC voxels were defined for use in Experiment 1. An anatomical mask of the occipitotemporal cortex was defined on the surface, with early visual regions (V1-V3) localized from the retinotopy protocol removed (left panel). Task-active voxels were defined from the contrast of all conditions > rest with $t > 2$ in either texform runs or original runs (middle panel). Active OTC was taken as the intersection of these two masks and was used for subsequent analyses. This procedure was carried out in each participant.

Experiment 1

A. Animacy organization

Whole brain group topographies



B. Object size organization

Whole brain group topographies

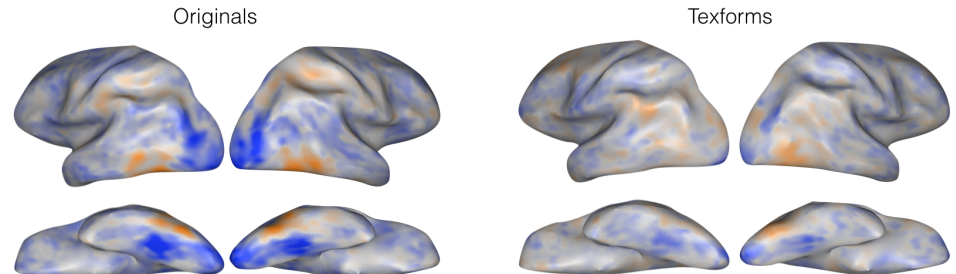
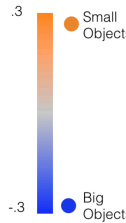


Figure S7. Whole brain group topographies for the animacy and object size distinctions, shown separately for originals (left panels) and texforms (right panels).

Experiment 1: Single subject topographies

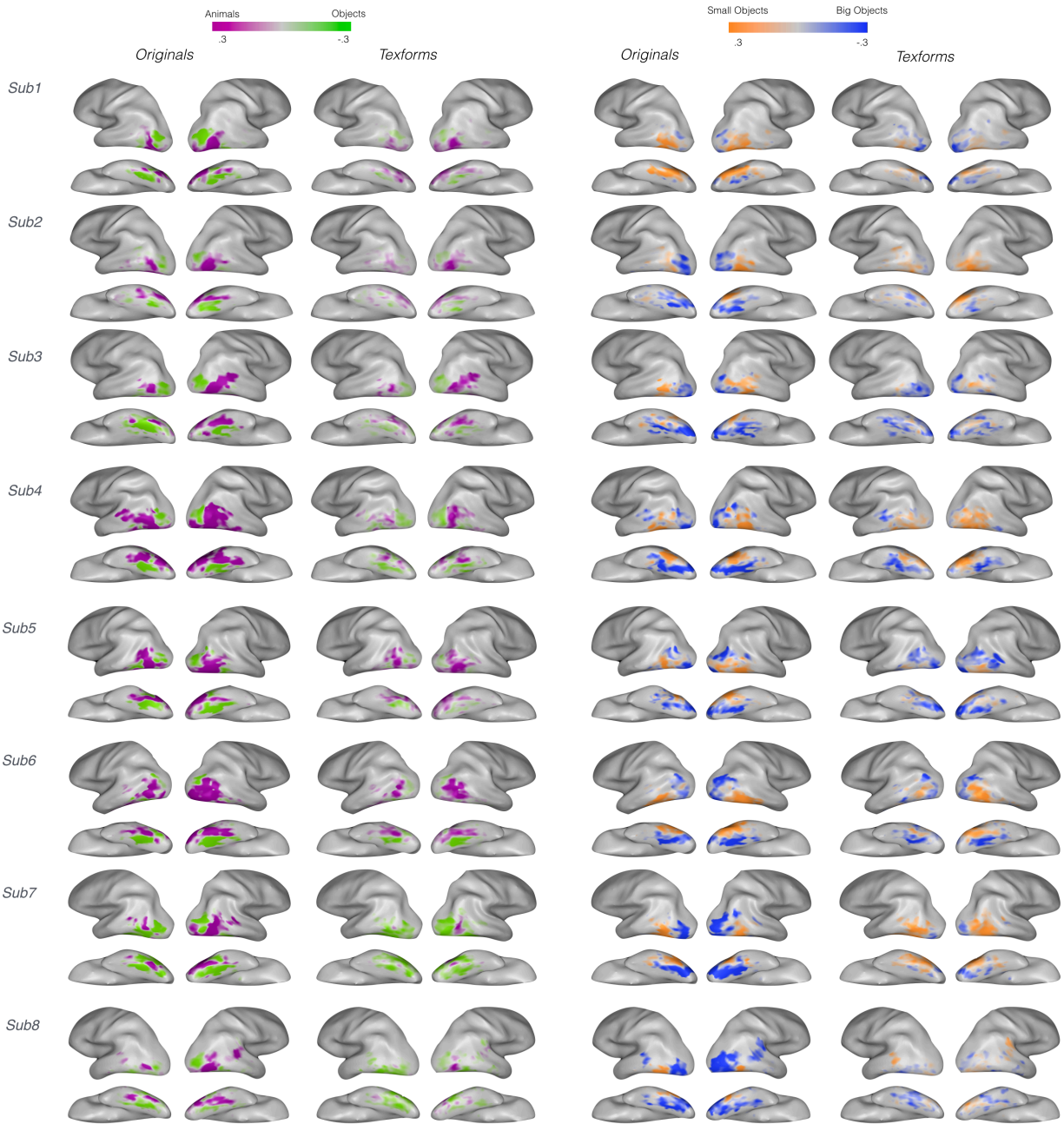
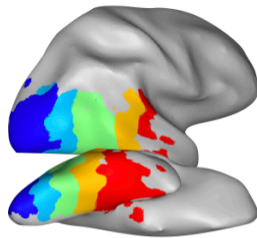


Figure S8. All single-subject topographies in both hemispheres for the animacy (left panels) and object size (right panel) distinctions, shown separately for originals and texforms. Preferences are shown within task-active occipito-temporal voxels.

A. Voxel sections

Active OTC
voxels
group data

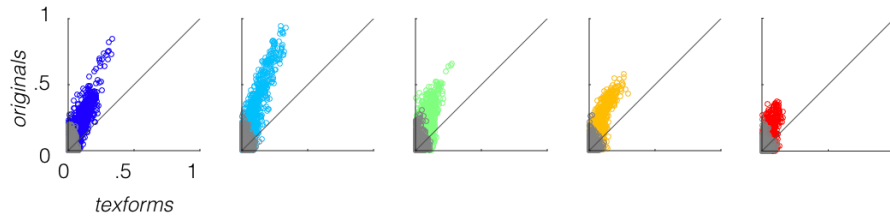
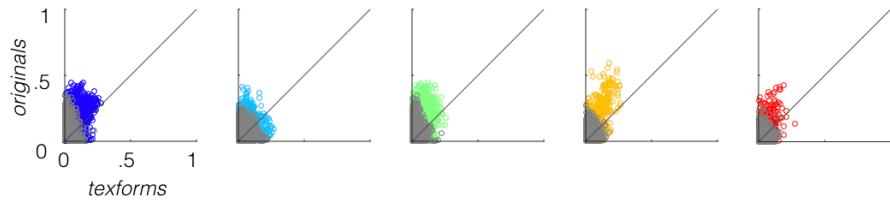
B. Animacy preference correlations**C. Object size preference correlations**

Figure S9. (A). Anatomical sections (shown here at the group level) from posterior to anterior in blue to red. (B) The animacy preferences elicited by texforms (x-axis) and by originals (y-axis) are plotted for each of the anatomical sections in the five subplots. Each point is a voxel. The x- and y-axes show strength of the animacy preference, computed as the absolute value of the difference between animal and object beta values. All points above the diagonal are voxels that show stronger animal/object preferences for original images than for texforms. Voxels where texforms and originals did not show the same preference are plotted in grey. (C) Object size preferences elicited by texforms (x-axis), and originals (y-axis) for each anatomical section.

Overall response differences between originals and texforms

Group data

Originals
> Texforms

.5

Texforms
> Originals

-.5

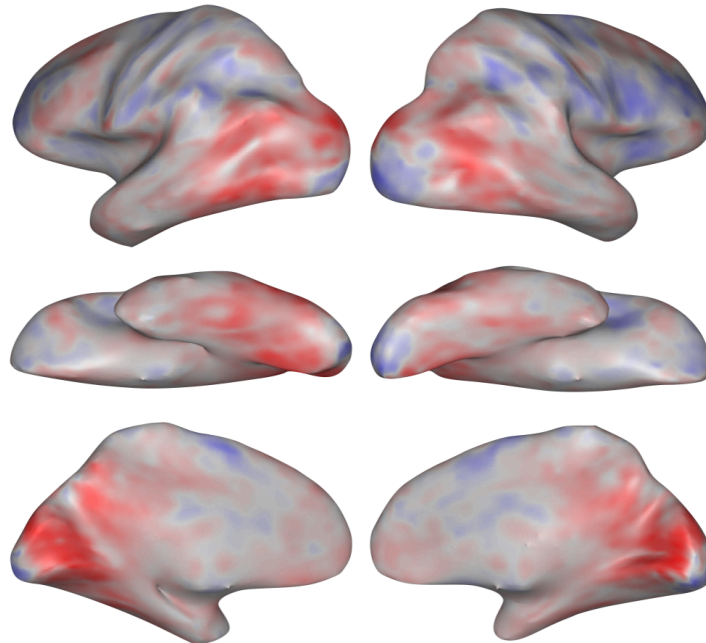


Figure S10. Overall activation differences between originals and texforms are shown at the group level. Voxels that showed stronger responses to originals are colored in red, and voxels that showed greater response differences to texforms are colored in blue.

Tripartite Organization

Group data

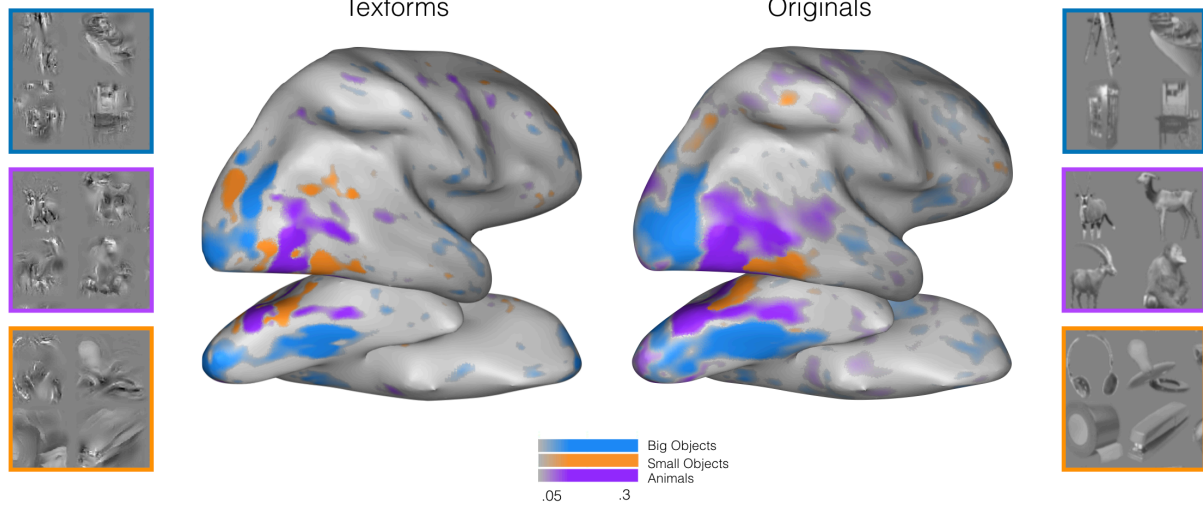


Figure S11. Comparison with Konkle & Caramazza, 2013 (5). (A) Group tripartite preference maps shown for both texforms (left) and originals (right) within task-active voxels.

4. Supplement to Experiment 2

Below, we show (a) maps of fixation stability for each participant (**Fig. S12**), (b) our procedure for defining location-tolerant voxels (**Fig. S13**), and (c) both group conjunction preference maps (**Fig. S14**) and all individual conjunction preference maps (**Fig. S15**).

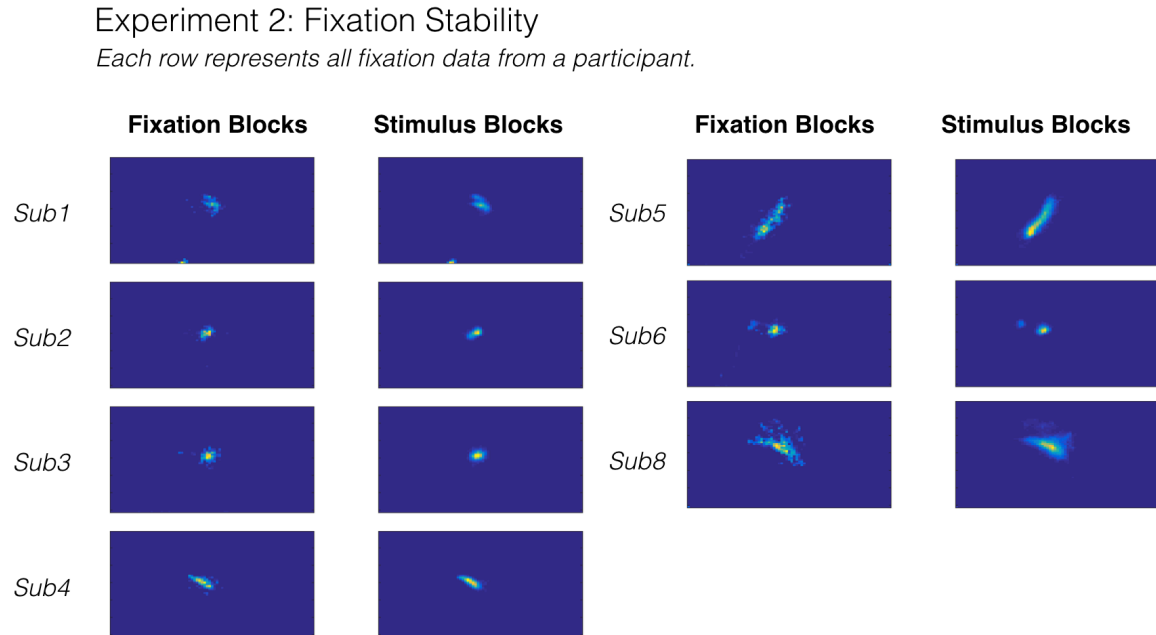


Figure S12. Fixation distributions are shown for each of the seven participants of Experiment 2 for whom we have eye-tracking data. The left column shows fixation distributions during time periods in which only a fixation dot was on the screen, and the right panel during time periods in which the stimuli were on screen at either an upper or lower visual location (in addition to the fixation dot). Note that while we were unable to obtain accurate calibrations for each participant, the deviations from fixation are highly similar between fixation and stimulus blocks. Thus, it is likely that this deviations from tight fixation reflect drift/noise in the calibration, rather than systematic looks towards the upper or lower visual field.

Defining location-tolerant voxels

Single subject example for original images

Defined individually for each subject, distinction, and image type

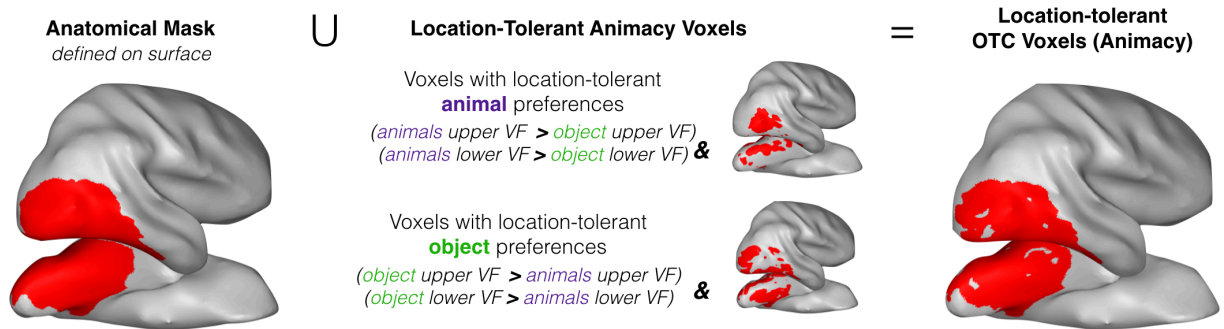
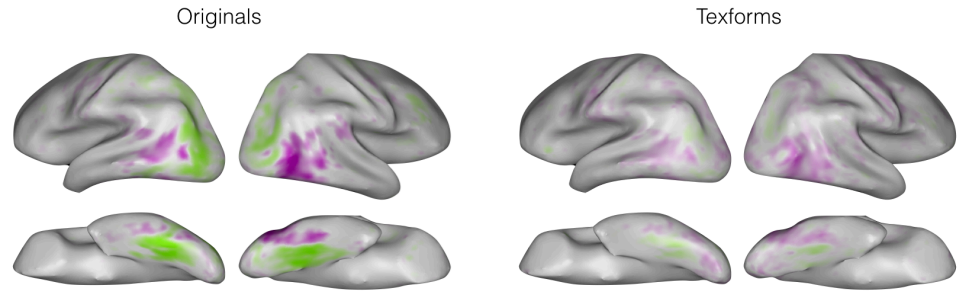


Figure S13. The procedure for defining location-tolerant voxels is shown for a single participant. First, an anatomical mask was defined on the surface for each participant to include occipitotemporal voxels and exclude early visual voxels (V1-V3) localized from a separate retinotopy protocol. Next, location-tolerant voxels were computed for each contrast (e.g., animals > objects and objects > animal) within this anatomical mask (middle panel). Finally, these two sets of location-tolerant voxels that prefer animals and objects, respectively, were fused together to create the final conjunction mask. The same procedure was followed for the object size distinction. These masks were computed separately for original images and texform images, in each participant.

Experiment 2

A. Animacy organization

Whole brain group topographies



B. Object size organization

Whole brain group topographies

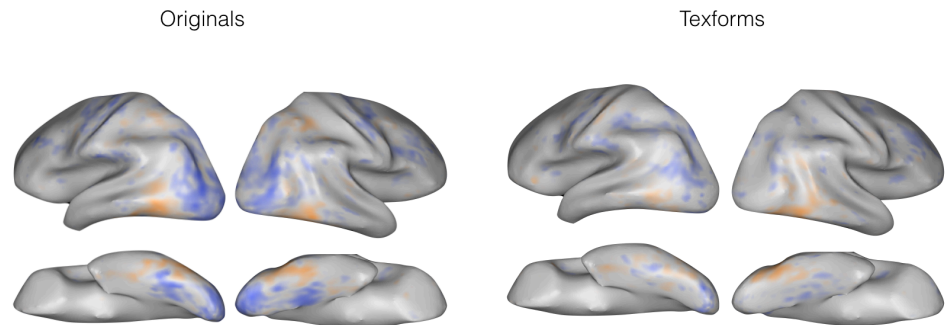
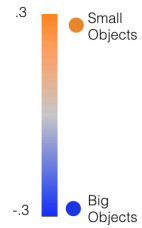


Figure S14. Group-level conjunction topographies for animacy (top) and object size (bottom) for texforms and originals. Preferences are shown within location-tolerant OTC voxels to originals (see Fig. S13).

Experiment 2: Single subject conjunction topographies

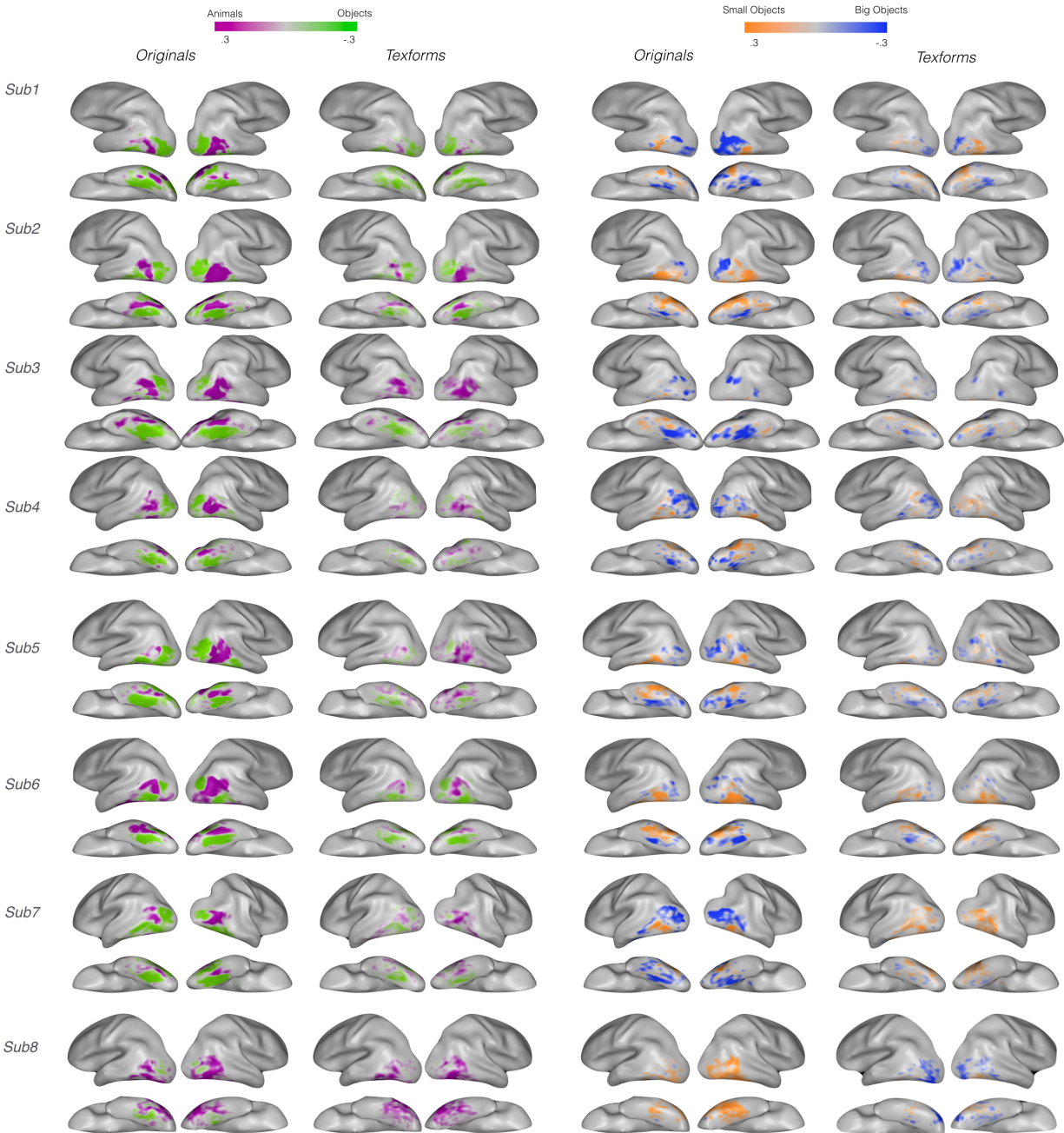


Figure S15. All single-subject conjunction topographies are shown in both hemispheres for the animacy (upper panel) and object size (lower panel) distinctions, shown separately for originals and texforms. Preferences are shown within location-tolerant OTC voxels to originals (see Fig. S13).

5. Supplement to Predictive Modeling Analyses

Below, we illustrate how reliable, task-active voxels were selected for the predictive modeling analyses (**Fig. S16**). We then report the results of additional modeling analyses along a posterior – anterior gradient of the ventral stream (**Fig. S17, S18**) and in early visual cortex (**Fig. S19**).

Defining reliable voxels in OTC

Single subject example - defined individually for each subject

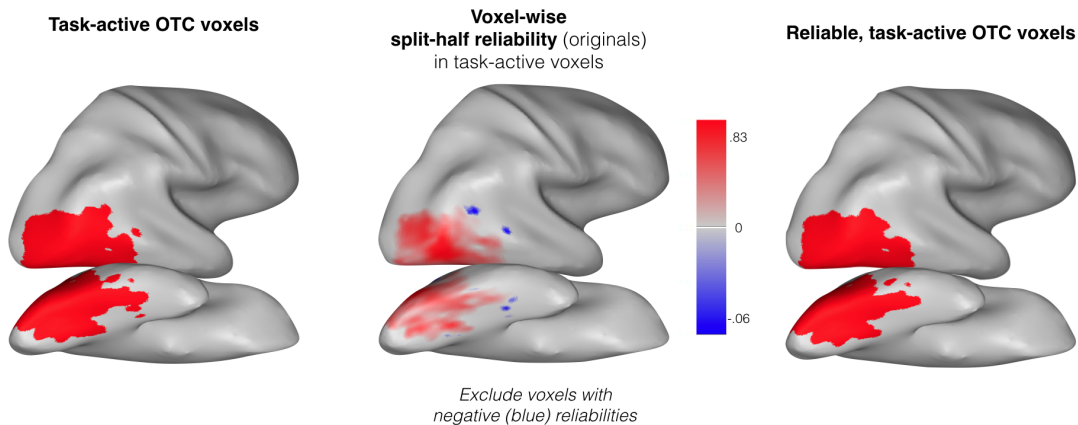
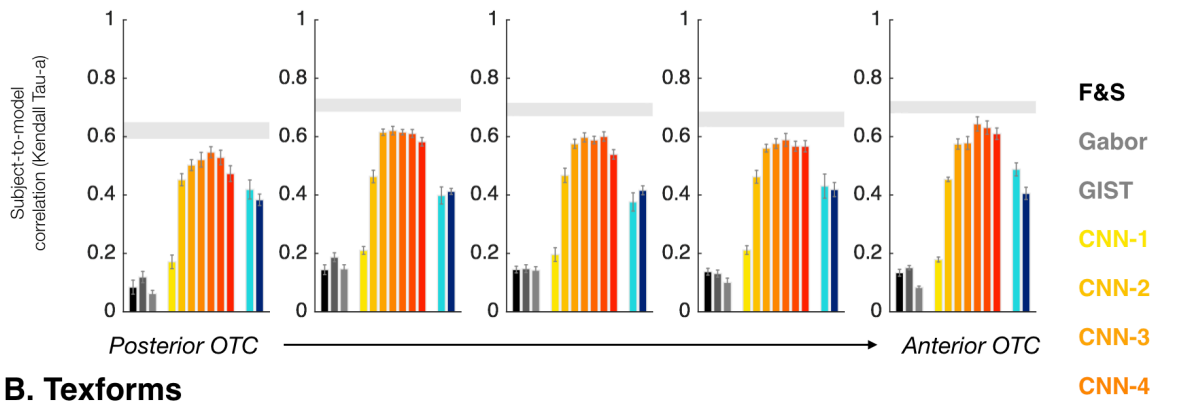


Figure S16. The procedure for reliable OTC voxels is shown for a single participant. First, we started with the task active voxel mask, as defined in **Figure S6**. Next, for each of these voxels, we extracted data from the condition-rich design (24 conditions), in odd and even runs in which original images were presented. The correlation between these two activation profiles was computed, and these voxel-wise split-half reliabilities are plotted in the middle panel. Voxels that are colored blue have slightly negative reliabilities. For the predictive modeling analysis, we excluded any voxel with a split-half reliability below zero. The right panel shows the final set of selected voxels for analysis for this participant. This voxel selection procedure was made a priori, with the motivation that it makes sense to only allow voxels that positively correlate with themselves in odd-even halves of the data to contribute to the final neural RDM. That being said, we also tested the impact of this choice in a post-hoc analysis, by repeating the analyses on the full set of active OTC voxels. The modeling results were extremely similar both qualitatively and quantitatively.

Model Performance, Anterior-to-Posterior Gradient

A. Originals



B. Textforms

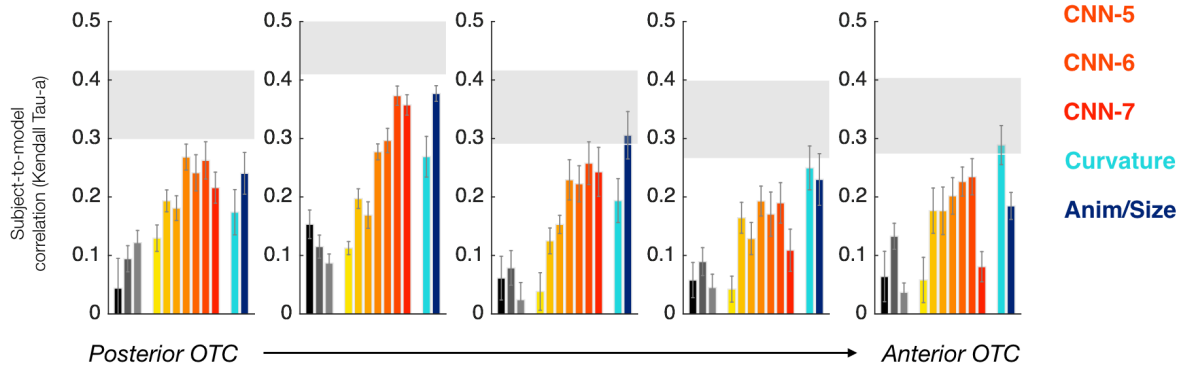
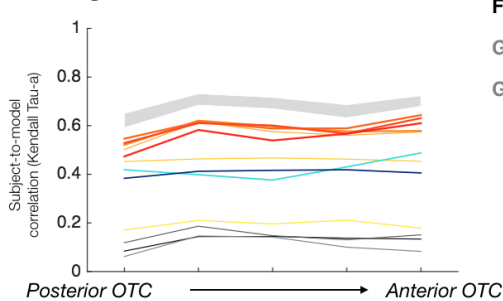


Figure S17. Predictive modeling results by anatomical sections of occipito-temporal cortex for originals (A) and textforms (B).

A. Originals



B. Textforms

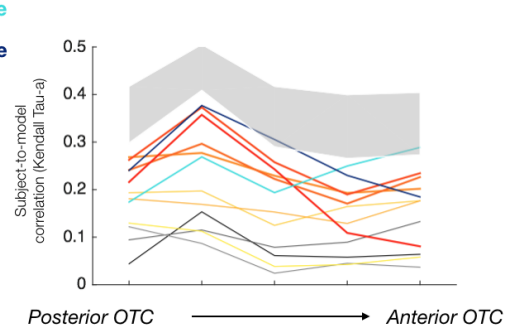


Figure S18. A summary of the results from Fig. S17 are shown, where the performance of each model across regions is plotted for both originals (A) and textforms (B).

Modeling in Early Visual Cortex (EVC). Given that we generated texforms using a texture synthesis model that explains variance in V2/V4 (6, 7), we explored which feature spaces explained variance in early visual cortex by applying the same analytic method. Consistent with prior work, we found that Gabor-based models explained the most variance in early visual cortex, whereas models based on perceptual properties (e.g., perceived curvature) or category-based models explained less variance.

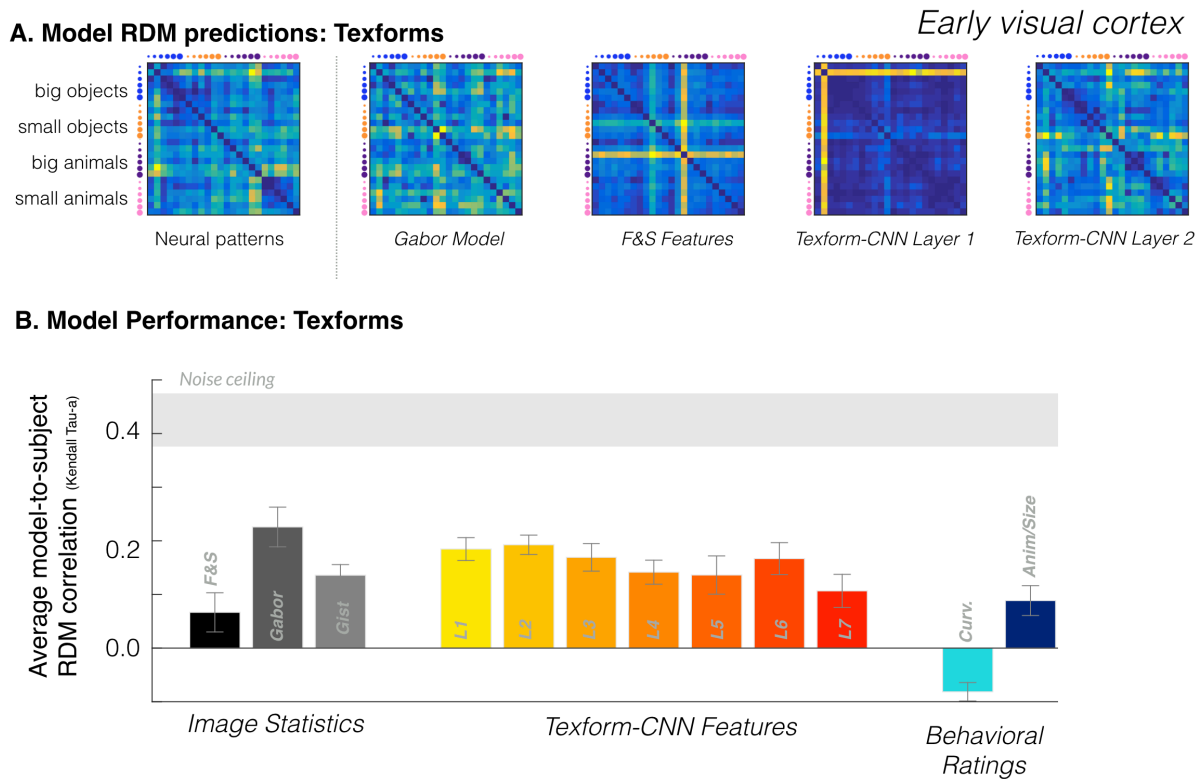


Figure S19. (A) Neural patterns in response to texforms in *early visual cortex* and predicted neural dissimilarities for selected models obtained through the same cross-validation procedure. (B) Average predicted model correlation (Kendall Tau- α) with individual subjects' neural patterns in EVC. Data is plotted with respect to the noise ceiling, shown in light gray.

References:

1. Cohen MA, Konkle T, Rhee JY, Nakayama K, Alvarez GA (2014) Processing multiple visual objects is limited by overlap in neural channels. *Proc Natl Acad Sci* 111(24):8955–8960.
2. Krizhevsky A, Sutskever I, Hinton G (2012) Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst*:1097–1105.
3. Freeman J, Simoncelli EP (2011) Metamers of the ventral stream. *Nat Neurosci* 14(9):1195–1201.
4. Willenbockel V, et al. (2010) Controlling low-level image properties: The SHINE toolbox. *Behav Res Methods* 42(3):671–684.
5. Konkle T, Caramazza A (2013) Tripartite organization of the ventral stream by animacy and object size. *J Neurosci* 33(25):10235–42.
6. Freeman J, Ziemba CM, Heeger DJ, Simoncelli EP, Movshon JA (2013) A functional and perceptual signature of the second visual area in primates. *Nat Neurosci* 16(7):974–81.
7. Okazawa G, Tajima S, Komatsu H (2015) Image statistics underlying natural texture selectivity of neurons in macaque V4. *Proc Natl Acad Sci U S A* 112(4):E351-60.