

RESEARCH ARTICLE

A data-driven approach to stimulus selection reveals an image-based representation of objects in high-level visual areas

David D. Coggan¹  | Afrodite Giannakopoulou¹ | Sanah Ali¹ | Burcu Goz¹ |
David M. Watson² | Tom Hartley^{1,3} | Daniel H. Baker^{1,3}  | Timothy J. Andrews¹ 

¹Department of Psychology, University of York, York, UK

²School of Psychology, The University of Nottingham, Nottingham, UK

³York Biomedical Research Institute, University of York, York, UK

Correspondence

Timothy J. Andrews, Department of Psychology, University of York, York, UK.
Email: timothy.andrews@york.ac.uk

Abstract

The ventral visual pathway is directly involved in the perception and recognition of objects. However, the extent to which the neural representation of objects in this region reflects low-level or high-level properties remains unresolved. A problem in resolving this issue is that only a small proportion of the objects experienced during natural viewing can be shown during a typical experiment. This can lead to an uneven sampling of objects that biases our understanding of how they are represented. To address this issue, we developed a data-driven approach to stimulus selection that involved describing a large number of objects in terms of their image properties. In the first experiment, clusters of objects were evenly selected from this multi-dimensional image space. Although the clusters did not have any consistent semantic features, each elicited a distinct pattern of neural response. In the second experiment, we asked whether high-level, category-selective patterns of response could be elicited by objects from other categories, but with similar image properties. Object clusters were selected based on the similarity of their image properties to objects from five different categories (bottle, chair, face, house, and shoe). The pattern of response to each metameric object cluster was similar to the pattern elicited by objects from the corresponding category. For example, the pattern for bottles was similar to the pattern for objects with similar image properties to bottles. In both experiments, the patterns of response were consistent across participants providing evidence for common organising principles. This study provides a more ecological approach to understanding the perceptual representations of objects and reveals the importance of image properties.

KEYWORDS

data-driven, fMRI, objects, ventral visual pathway

1 | INTRODUCTION

Visual areas involved in object perception form a ventral processing pathway that projects from the occipital toward the temporal lobe (Milner & Goodale, 1995; Ungerleider & Mishkin, 1982). Lesions to

different regions of the ventral visual pathway can produce selective deficits in the perception and recognition of different objects (McNeil & Warrington, 1993; Moscovitch, Winocur, & Behrmann, 1997). Consistent with these neuropsychological reports, neuroimaging studies have revealed discrete regions that are specialised for

different categories of objects (Cohen et al., 2000; Downing, Jiang, Shuman, & Kanwisher, 2001; Epstein & Kanwisher, 1998; Kanwisher, McDermott, & Chun, 1997; McCarthy, Puce, Gore, & Truett, 1997). Although specialised regions have only been reported for a limited number of categories (Downing, Chan, Peelen, Dodds, & Kanwisher, 2006; Vul, Lashkari, Hsieh, Golland, & Kanwisher, 2012), multivariate analyses have shown that the spatial pattern of neural response is able to discriminate a greater range of object categories (Haxby et al., 2001; Kriegeskorte et al., 2008). These distributed patterns of response shown in these multivariate studies are thought to reflect a topographic organisation of objects that is analogous to that found in early visual areas, where the topography is tightly linked to basic visual properties (Bonhoeffer & Grinvald, 1991; Engel et al., 1994; Hubel & Wiesel, 1968; Wandell & Winawer, 2011). However, it remains unclear what organising principles might underpin the topographic representations of objects in higher visual areas.

Understanding these organising principles is challenging, because high-level and low-level properties often covary in natural objects (Malcolm et al., 2016; Rice, Watson, Hartley, & Andrews, 2014). Patterns of response in higher-visual areas of the ventral visual pathway have been linked to higher-level properties of objects, such as category (Connolly et al., 2012; Haxby et al., 2001), animacy (Kriegeskorte et al., 2008), semantics (Naselaris, Prenger, Kay, Oliver, & Gallant, 2009) and real-world size (Konkle & Oliva, 2012). However, it remains unclear how these representations emerge from the image-based representations found in early visual areas. One possibility is that the patterns of response in high-level visual areas reflect an underlying representation that is based on more fundamental properties of the stimulus (Andrews, Watson, Rice, & Hartley, 2015). Electrophysiological studies in

nonhuman primates also suggest that the topography of higher visual areas is based on a continuous mapping of image features rather than discrete object representations (Tanaka, 1996). However, the sparse nature of these recordings makes it difficult to determine with certainty the critical dimensions along which information is represented. Neuroimaging studies have also shown that differences in the visual properties of objects can explain a significant amount of the variance in high-level regions of visual cortex (Coggan et al., 2019; Levy et al., 2001; Rice et al., 2014; Nasr et al., 2014; Watson, Hartley, & Andrews, 2014; Watson, Young, & Andrews, 2016; Sormaz et al., 2016). For example, category-selective patterns of response are still evident when images have been scrambled in a way that preserves some of their visual properties, but removes their semantic properties (Andrews et al., 2010; Coggan, Baker, & Andrews, 2016; Coggan, Liu, Baker, & Andrews, 2016; Long, Yu, & Konkle, 2018; Watson, Andrews, & Hartley, 2017). A number of neuroimaging studies have directly compared the influence of low-level and high-level properties on patterns of response to objects (Kriegeskorte et al., 2008; Lescroart and Gallant, 2019; Naselaris et al., 2009; Clarke & Tyler, 2014; Bracci & Op de Beeck, 2016; Proklova, Kaiser, & Peelen, 2016; Jozwik, Kriegeskorte, & Mur, 2016). It is becoming clear from these studies that the representation across the ventral stream reflects both high-level and low-level representations. Nevertheless, these studies vary in the extent to which they support the importance of low-level and high-level properties.

Key to understanding how objects are represented in the brain is the ability to uniformly sample the vast number of objects we encounter during a lifetime of natural viewing. In a typical neuroimaging experiment, only a finite number of images can be presented. This can lead to experimental designs that compare responses to relatively small numbers of

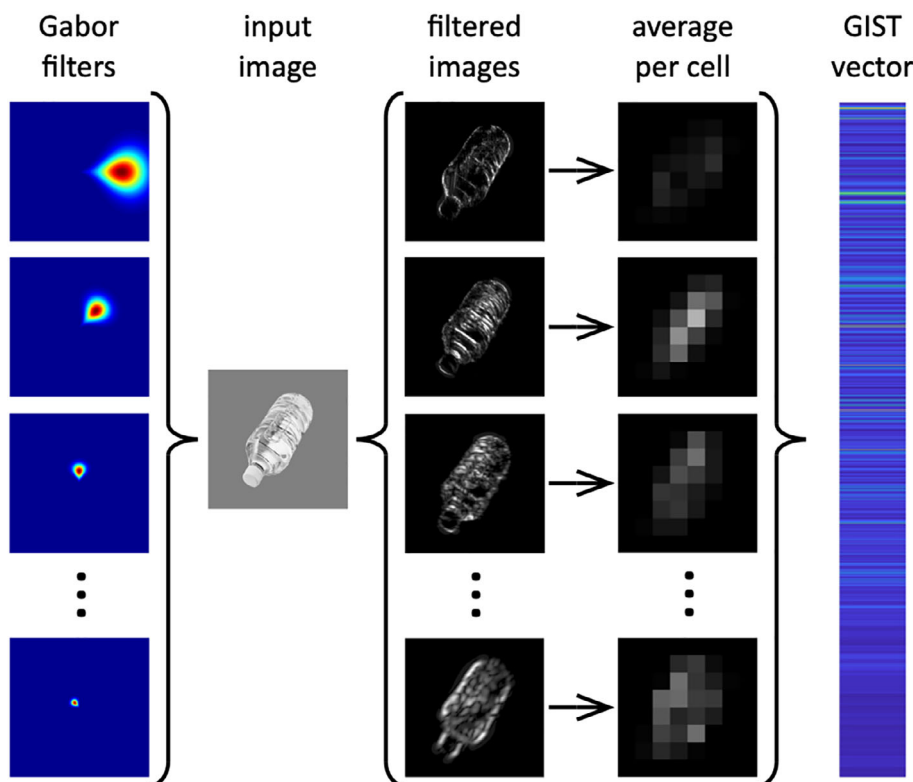


FIGURE 1 GIST descriptor (Oliva & Torralba, 2001) applied to an example image. The 64 Gabor filters (shown here in Fourier space) were constructed across factorial combinations of eight spatial frequencies and eight orientations. Each filter was applied to the image in turn, resulting in 64 filtered images. Each filtered image was then windowed into an 8×8 grid and pixel intensities within each window were averaged. These were then concatenated into a single vector of 4,096 values that describes the spatial frequency and orientation information present across the image [Color figure can be viewed at wileyonlinelibrary.com]

objects from experimenter-defined stimulus conditions, which makes it difficult to disentangle the subjective manipulation of higher-level dimensions of the stimulus from those driven by correlated lower-level dimensions. To understand how the neural representation of objects might emerge, it is necessary to develop methods to sample objects in a more objective, ecologically valid way and then determine how they affect patterns of response across visual cortex. In this study, we used a data-driven approach to select stimuli (Watson et al., 2017), and measured neural responses to these images using fMRI. Images from a large object database were described in terms of their image properties and clustering algorithms were used to evenly sample clusters of objects from this image space. Our rationale is that these object clusters will provide a good first approximation to the diversity of objects that an individual will be exposed to during a lifetime of natural viewing and avoids the need to impose any additional higher-level constraints on stimulus selection.

Our aim was to determine whether clusters of objects defined purely by their image properties would generate distinct patterns of response in the ventral stream in the same way as is typically observed for subjectively defined image categories. If this were the case, this would suggest that the neural representation of objects was to some degree based on their underlying image properties. As a further test of the role of low-level properties in the organisation of the ventral visual pathway, we asked whether high-level representations of category could be explained by low-level properties. Clusters of objects were selected from the database based on their similarity to objects from commonly used object categories (bottle, chair, face, house, and shoe). Our aim was to determine whether these object clusters would elicit similar patterns of response to objects defined by object category. For example, is the pattern of response to chairs similar to the pattern of response to objects that have similar image properties to chairs? Patterns of neural response were compared across participants to determine inter-subject consistency (see Coggan et al., 2017; Flack et al., 2015; Rice et al., 2014; Watson et al., 2014; Weibert, Flack, Young, & Andrews, 2018). This is important as it tests whether common organising principles underpin the topography of the response across participants.

2 | MATERIALS AND METHODS

2.1 | Participants

Twenty-one participants took part in Experiment 1 (11 male, mean age = 23, $SD = 2.7$ years). Twenty-eight participants took part in Experiment 2, with data from three participants removed for excessive movement during the scan (final sample: 16 male, mean age = 25.7, $SD = 7.0$ years). Sample size was based on previous studies using similar designs (Coggan, Baker, & Andrews, 2016; Watson et al., 2017). All participants were right-handed, had normal or corrected-to-normal vision and no history of mental illness. Each gave their informed, written consent and the study was approved by the York Neuroimaging Centre (YNiC) Ethics Committee and adhered to the original wording of the Declaration of Helsinki. Stimuli were back-projected onto a custom in-bore acrylic screen and viewed via a mirror placed above the subject's head. Viewing distance was approximately 57 cm. All

objects were presented within a $15 \times 15^\circ$ frame, though the objects themselves subtended a smaller visual angle than this. Stimuli were presented using Psychopy (Peirce, 2007).

2.2 | Image properties

In order to obtain a realistic range of real-world objects, we used all images contained in the Bank of Standardised Stimuli (Brodeur, Dionne-Dostie, Montreuil, & Lepage, 2010) as this comprises a large and diverse range of objects (2,761 objects at the time of selection) taken from a range of categories that includes: food, tool, musical instruments, vehicles, weapons, animals, body parts, buildings, clothing, electronic devices, furniture, games, jewellery, kitchen utensils, medical instruments, sports items, bathroom items, stationary (see table 2 of Brodeur et al., 2010). Images were converted to greyscale and then measured with the GIST descriptor (Oliva & Torralba, 2001), which describes the spatial frequency and orientation information present at different spatial locations across the image as a numerical vector (Figure 1). We configured the descriptor to measure the energy at eight spatial frequencies across eight orientations and 64 spatial subdivisions (8×8) of the image, resulting in a vector of 4,096 values that described each image. GIST vectors were then normalised by first scaling each component of the vectors to sum to one across images, and second by scaling each vector to have a magnitude of one. Our motivation for using GIST is primarily that we have used this method with success in a number of previous studies in this area (Rice et al., 2014; Watson et al., 2014; Watson et al., 2016; Watson et al., 2017; Weibert et al., 2018). An advantage of this method is that it makes minimal assumptions about the way image properties are represented.

2.3 | Experiment 1

2.3.1 | Image selection

A k-means clustering algorithm was used to evenly select clusters of objects with similar image properties from different regions of the object space defined by the 4,096 dimensions of the GIST descriptor. Attempting to apply clustering algorithms in a high-dimensional space can be problematic and can result in slow or unreliable clustering leading to the selection of atypical or outlier objects (Bellman, 1961). So we first reduced the dimensionality using principal components analysis (PCA). We used the first 20 principal components, which explained 58.0% of the variance of the original components (see Watson et al., 2017). We applied the k-Means clustering algorithm ($k = 10$; Euclidean distance metric) to identify 10 centroids within this space. We then selected the 24 images nearest the centroid of each cluster as measured by Euclidean distance, such that images within a cluster have similar visual properties to one another.

This process of image selection is shown in Figure 2. The choice of cluster number and the number of images were chosen to fit within the constraints of a neuroimaging study, rather than to fit any assumptions about the structure of the image space. We would predict that a similar pattern of results would be evident with a different value of k , provided there was sufficient power in fMRI analysis. PCA

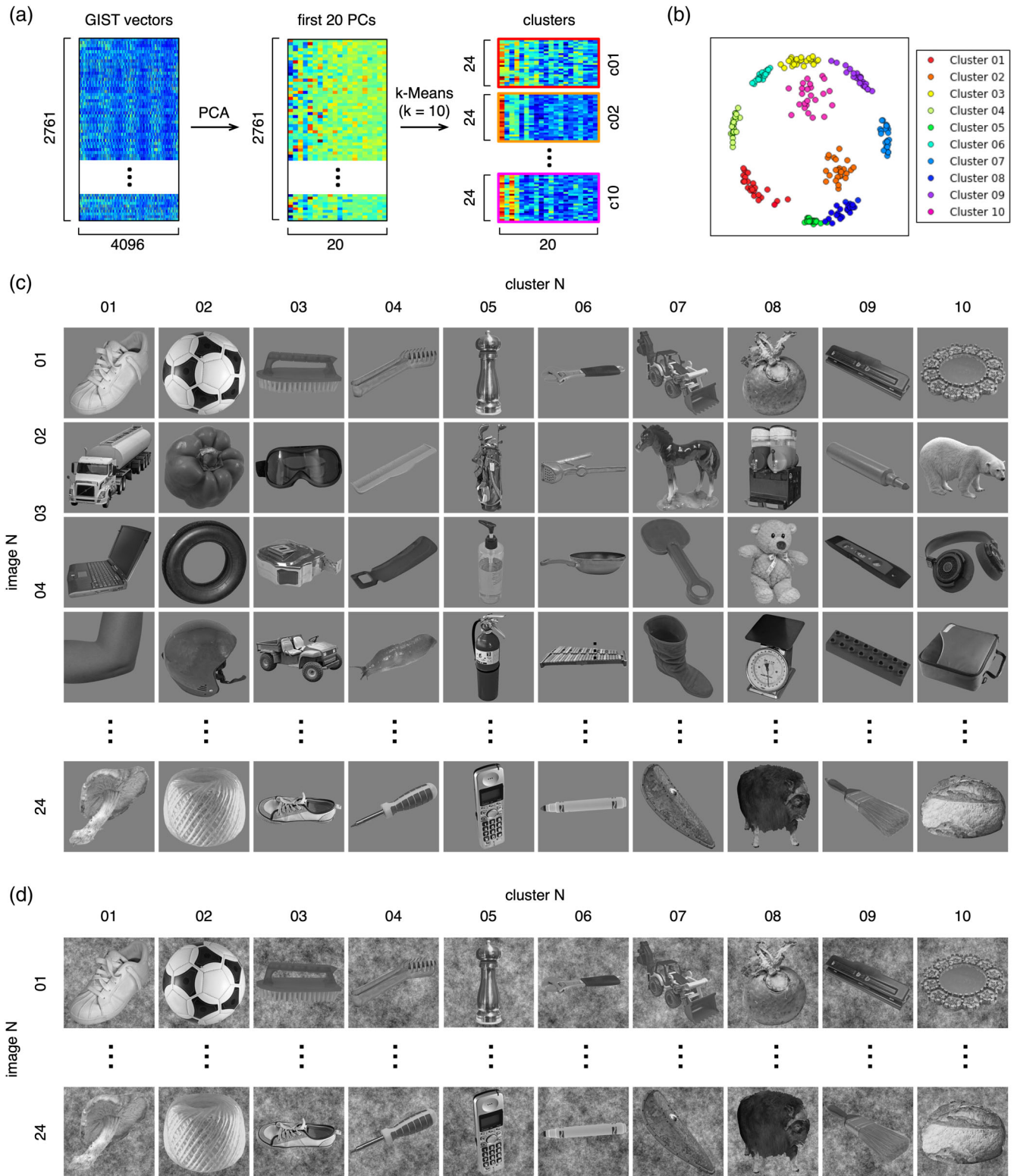


FIGURE 2 Data-driven image selection in Experiment 1. (a) GIST descriptions were generated for each image in the BOSS database (see Figure 1). PCA was used to reduce the dimensionality of the data, with the first 20 PCs selected. Ten distinct clusters within this feature space were then defined through k-means clustering, with the 24 most proximate images to each cluster centroid selected to represent each cluster. (b) Multidimensional scaling plot approximating the locations of the selected images within the feature space. (c) Examples of stimuli from each of the 10 clusters on a uniform, mid-grey background ('untextured' condition). (d) The same stimulus set was also superimposed on a pink noise background ('textured' condition) [Color figure can be viewed at wileyonlinelibrary.com]

and k-Means algorithms were implemented using the Python Scikit-learn toolbox (Pedregosa et al., 2011). Multidimensional scaling was also used to visualise the locations of images in each cluster in a 2D approximation of the principal component feature-space (Figure 2b). The mean distance between exemplars and the centroid in this plot was similar across clusters (01:0.26 \pm 0.13; 02:0.14 \pm 0.07; 03:0.19 \pm 0.07; 04:0.13 \pm 0.07; 05:0.07 \pm 0.04; 06:0.09 \pm 0.05; 07:0.18 \pm 0.07; 08:0.22 \pm 0.08; 09:0.17 0.09; 10:0.31 \pm 0.11; AU \pm SD).

Two versions of this stimulus set were then created by applying a uniform, mid-grey background (untextured condition, Figure 2c) and a unique, pink noise (1/f) background (textured condition, Figure 2d) to each of the 240 images. All images are shown in Figure S1. The rationale for the textured background was that images rarely appear in isolation, so we were interested to know if a similar pattern of response was evident when images were presented on uniform compared to a textured background. The textured backgrounds were designed to provide visual stimulation across the extent of the image, which emulates the 1/f amplitude energy distribution found in most natural images, without conveying any confounding visual or semantic information.

2.3.2 | Design and procedure

The fMRI experiment consisted of two scans. Images with untextured and textured backgrounds were presented in different scans, the order of which was counterbalanced across subjects. In each scan, objects from the 10 clusters were presented in 6 s blocks. In each block, six objects from the same cluster were presented individually for 800 ms, with a 200 ms inter-stimulus-interval. This was followed by a fixation cross lasting 9 s. Images from each cluster were shown four times, giving a total of 40 blocks. The order of the blocks was randomised. Participants performed a task whilst viewing images, designed to maintain attention for the duration of the scan. The task consisted of pressing a button on a response box whenever a red dot appeared on an image. Red dots were randomly placed on 40 of the 240 images presented throughout the scan.

2.4 | Experiment 2

2.4.1 | Image selection

The procedure for image selection in Experiment 2 is shown in Figure 3. Image clusters were selected based on their image similarity (GIST vector) to images from five categories that are commonly used to test the response of objects in the ventral visual pathway (Haxby et al., 2001; Rice et al., 2014). The five categories were bottles, chairs, faces, houses, and shoes. The GIST descriptor was applied to 36 images from each object category that have been used in previous experiments (Coggan, Liu, et al., 2016; Rice et al., 2014; Watson et al., 2016). From these images, an average GIST descriptor was calculated for each category. The average GIST vector for each category was correlated with the GIST vectors from each image in the BOSS database. For each of the five categories, all objects in the BOSS database were ranked based on correlation. For each category, 36 images with

the highest correlation values were selected, excluding those images that had an obvious semantic relationship with any of the five original object categories. All images are shown in Figure S2.

2.4.2 | Design and procedure

The fMRI experiment had 10 conditions. Five conditions contained images from five object categories (bottle, chair, face, house, and shoe). The other five conditions had images with similar image properties to bottles, chairs, faces, houses, and shoes. In each scan, objects from the 10 conditions were presented in 6 s blocks. In each block, six objects from the same cluster were presented individually for 800 ms, with a 200 ms inter-stimulus-interval. This was followed by a fixation cross lasting 9 s. Images from each cluster were shown six times, giving a total of 60 blocks. The order of the blocks was randomized. Again, participants performed an orthogonal task designed to maintain attention. For this experiment, subjects had to respond whenever the fixation cross changed from black to green, which occurred on average once per block, randomly distributed throughout the scan duration.

2.5 | Data acquisition

fMRI data for both experiments were acquired with a General Electric 3 T HD Excite MRI scanner at YNIC at the University of York, fitted with an eight-channel, phased-array, head-dedicated gradient insert coil tuned to 127.4 MHz. A gradient-echo echo-planar imaging (EPI) sequence was used to collect data from 38 contiguous axial slices (TR = 3,000 ms, TE = 32.7 ms, FOV = 288 \times 288 mm, matrix size = 128 \times 128, voxel dimensions = 2.25 \times 2.25 \times 3 mm, flip angle = 90°). The fMRI data were analysed with FEAT v5.98 (<http://www.fmrib.ox.ac.uk/fsl>). In all scans, the initial 9 s of data were removed to reduce the effects of magnetic saturation. Motion correction (MCFLIRT, FSL) and slice-timing correction were applied, followed by temporal high-pass filtering (Gaussian-weighted least-squares straight line fitting, sigma = 50 s). Gaussian spatial smoothing was applied at 6 mm FWHM. Parameter estimates were generated for each cluster by regressing the hemodynamic response of each voxel against a box-car function convolved with a single-gamma HRF. Functional data were first registered to a low-resolution T1-anatomical image oriented in the same plane as the EPI (TR = 2.5 s, TE = 9.98 ms, FOV = 288 \times 288 mm, matrix size = 512 \times 512, voxel dimensions = 0.56 \times 0.56 \times 3 mm, flip angle = 90°), then to a high-resolution T1-anatomical image (TR = 7.96 ms, TE = 3.05 ms, FOV = 290 \times 290 mm, matrix size = 256 \times 256, voxel dimensions = 1.13 \times 1.13 \times 1 mm, flip angle = 20°) and finally onto the standard MNI brain (ICBM152).

2.6 | Regions of interest

The ventral stream regions of interest (ROIs) is shown in Figure 4. To construct a mask of the ventral visual pathway, we selected a series of anatomical ROIs from the Harvard-Oxford cortical atlas based on the physical limits of ventral temporal cortex described by Grill-Spector and Weiner (Grill-Spector & Weiner, 2014). Specifically, these regions were:

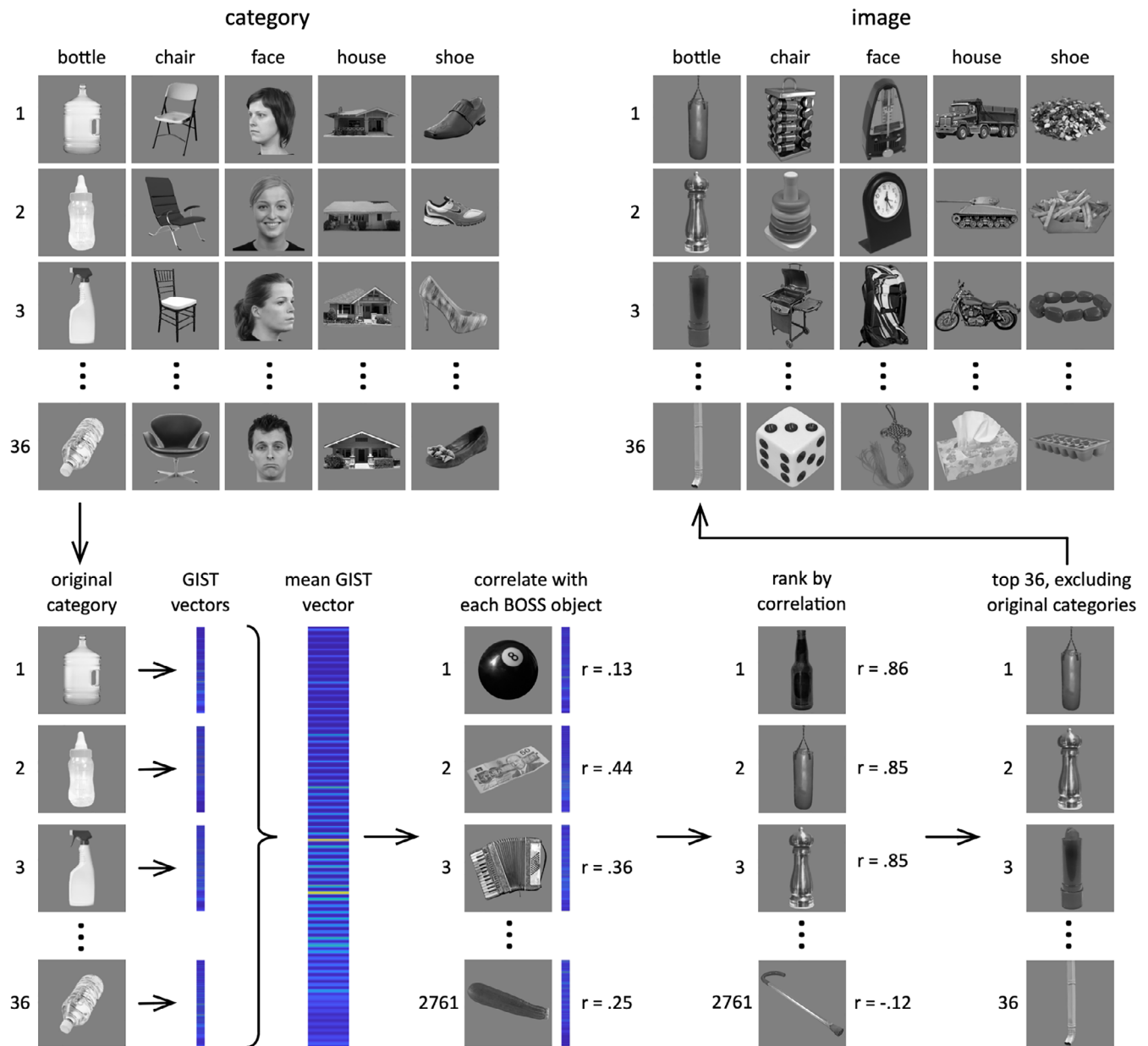


FIGURE 3 Image selection in Experiment 2. For each original category, 36 images were selected. Each image was then described by the GIST descriptor, as shown in Figure 1. GIST vectors were averaged across the 36 images to give a single vector for the category. This was then compared through correlation to GIST vectors generated for each object image in the Bank of Standardized Stimuli (BOSS). BOSS images were then ranked by their correlation coefficient and the top 36 images were selected. BOSS images containing objects from or closely related to any of the original categories were excluded. This resulted in 36 images with similar GIST vectors to the original category [Color figure can be viewed at wileyonlinelibrary.com]

inferior temporal gyrus (temporo-occipital portion), temporal-occipital fusiform cortex, occipital fusiform gyrus, and lingual gyrus.

2.7 | Multi-voxel pattern analysis

The reliability of patterns of neural response to each condition was tested using a leave-one-participant-out (LOPO) cross-validation paradigm (Poldrack, Halchenko, & Hanson, 2009; Rice et al., 2014), which allows us to measure the consistency of the pattern of response across participants.

Parameter estimates to each condition were normalised by subtracting the mean response per voxel per participant across all categories. These data were then submitted to a correlation-based multi-voxel pattern analyses (MVPA, Hanson, Matsuka, & Haxby, 2004; Haxby et al., 2001) implemented using the PyMVPA toolbox (Hanke et al., 2009). For each unique combination of conditions, the LOPO analysis compares the patterns of response in each participant with a corresponding group parameter estimate determined using a higher-level analysis of the remaining participants. This was repeated for each participant.

The correlation coefficients were then used to populate a representational similarity matrix, which shows the relative similarity of patterns of response to different object clusters. A Fisher's Z-transformation was then applied to the correlations prior to further statistical analysis. To determine whether there were reliable patterns of response to each object cluster, the within-cluster correlations (e.g., cluster 1 vs. cluster 1) were compared to the relevant between-cluster correlations (e.g., cluster 1 vs. cluster 2, cluster 1 vs. cluster 3, etc.).

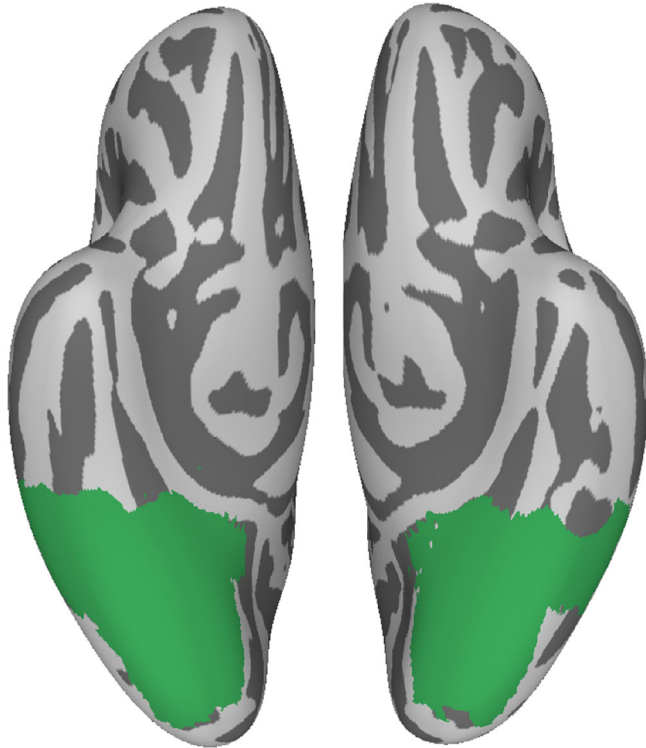


FIGURE 4 Ventral stream mask, projected onto a ventral view of inflated cortex [Color figure can be viewed at wileyonlinelibrary.com]

2.8 | Semantic similarity analysis

To generate a model of the similarity between object clusters based on semantic properties, we used WordNet—a lexical database of English (Miller, 1995). WordNet represents conceptual relations amongst nouns, including hyponymy (super- and subordinate categorical relations; e.g., between 'chair' and 'furniture') and meronymy (part-whole relations; e.g., between 'leg' and 'chair'), in a hierarchical taxonomy. The semantic similarity between a word pair can be estimated through their proximity in the taxonomy. This is illustrated in Figure 5a. We identified nouns in WordNet that corresponded to each image in our stimulus set, and generated semantic similarity estimates between all pairwise combinations. These estimates were then averaged for each pairwise combination of clusters to generate the similarity matrices shown in Figure 5b,c for Experiments 1 and 2, respectively. In Experiment 1 (Figure 5b), the semantic similarity was not significantly different for objects within an image cluster compared to objects in different object clusters ($t[9] = 1.70$, Cohen's $d = 0.54$, $p = .123$). In Experiment 2 (Figure 5c), not surprisingly, the semantic properties of objects chosen based on their category were more similar to each other than objects from different object categories ($t[9] = 113.81$, Cohen's $d = 43.25$, $p < .001$). However, semantic similarity was not significantly greater for objects within corresponding metameric clusters defined by image properties compared to objects from different metameric clusters ($t[5.15] = 113.81$, Cohen's $d = 1.04$, $p = .34$). There was also no semantic similarity between object clusters defined by category and metameric object clusters defined by image properties ($t[6.79] = 0.29$, Cohen's $d = 0.13$, $p = .78$). Together, these results show that object clusters defined by image properties did not have any consistent semantic properties.

3 | RESULTS

3.1 | Experiment 1

In Experiment 1, we measured the response to object clusters defined only by their image properties (see Figure 2). Figure 6a and Figure S3

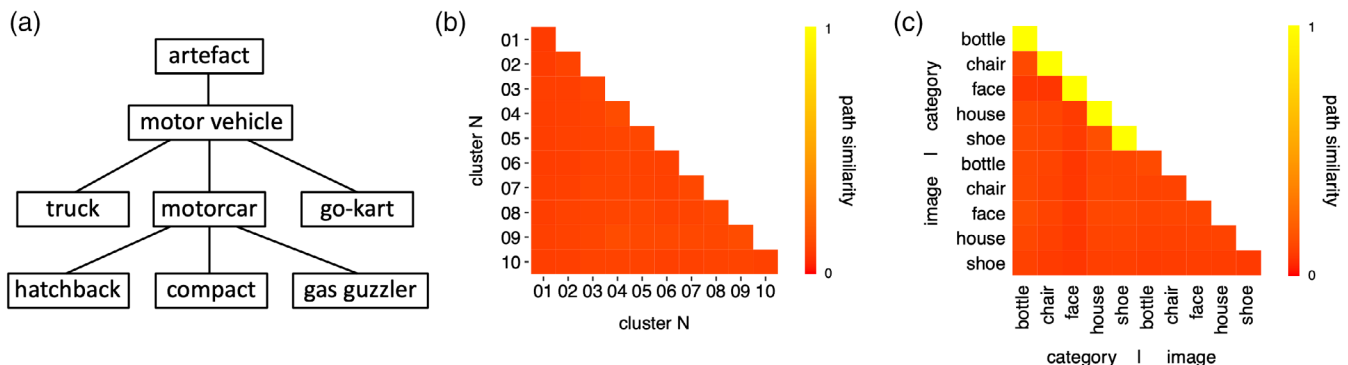


FIGURE 5 Semantic similarity analysis. (a) simplified illustration of Wordnet taxonomy structure. Semantic similarity is estimated by measuring the shortest number of connections between two words. For instance, 'hatchback' and 'compact' are separated by two connections, whereas 'hatchback' and 'truck' are separated by three. The former pair would therefore be estimated as more semantically similar than the latter. Matrices in b and c show average semantic similarity between conditions in Experiments 1 and 2, respectively. There was no consistency in the semantic properties of the images chosen based on their image properties [Color figure can be viewed at wileyonlinelibrary.com]

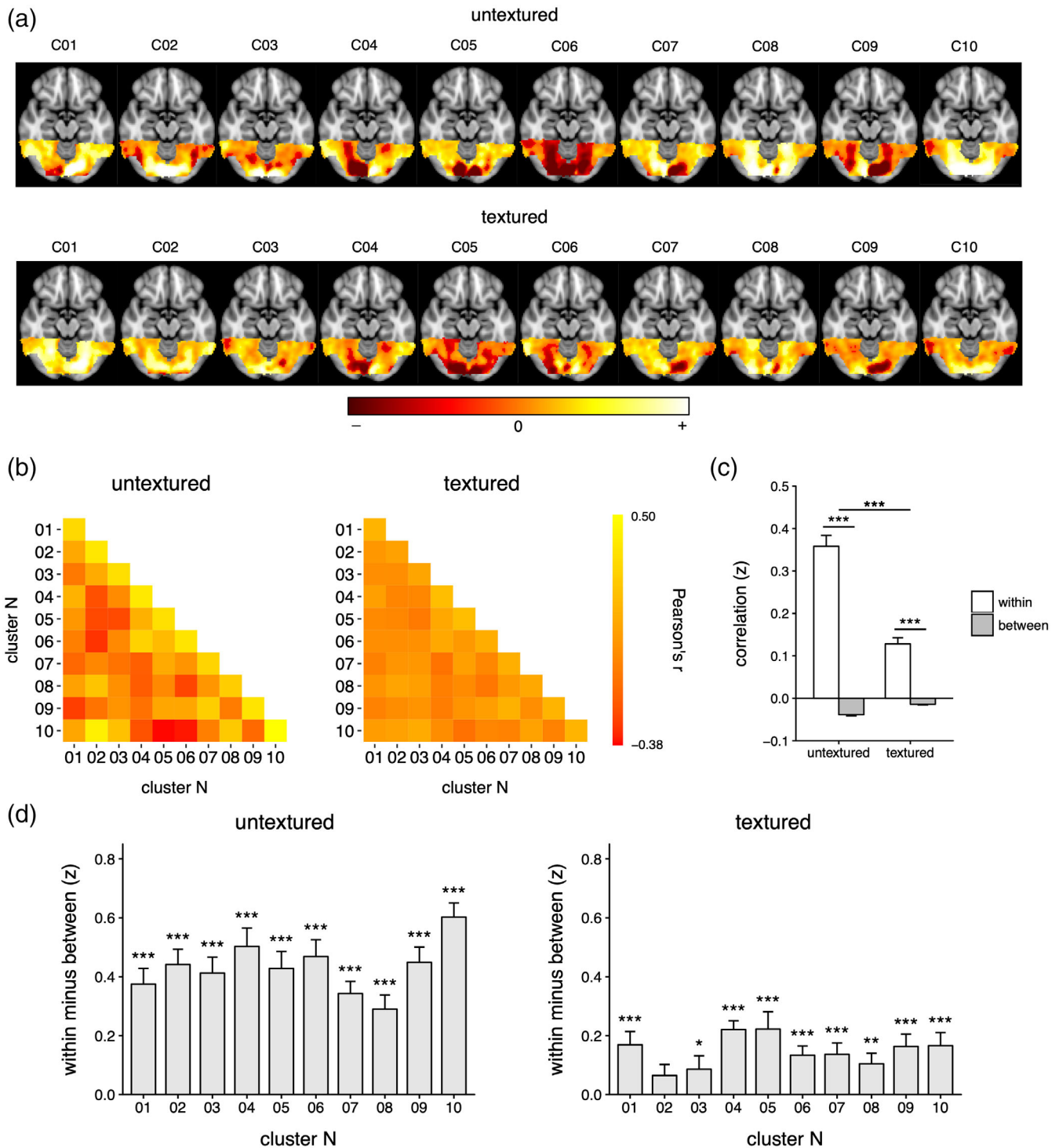


FIGURE 6 MVPA for Experiment 1. (a) Patterns of neural response in the ventral visual pathway to different object clusters with untextured and textured backgrounds. Patterns of response were normalised for each background type by subtracting the voxel-wise mean response across all 10 clusters from the response to each cluster. Axial slices are located at $z = -16$ (ICBM-MNI 152). (b) Group mean neural matrices showing correlations between neural responses within and between the different object clusters for untextured and textured conditions. Despite the difference in magnitude of the correlations in the untextured and textured matrices, there was a strong correlation between them ($r = 0.89$, $p < .001$; between-cluster correlations only). (c) Bar plot showing higher within-cluster compared to between-cluster correlations (Z-transformed) for textured and untextured conditions. (d) Bar plots showing the difference in within- and between-cluster correlations (Z-transformed) for each cluster. For all bar plots, error bars represent ± 1 SE of the mean. A similar pattern of results was evident when the data were analysed using a permutation test (Table S1). $*p < .05$, $**p < .01$, $***p < .001$ [Color figure can be viewed at wileyonlinelibrary.com]

show the average pattern of response to each of the object clusters in the ventral visual pathway ROI. Figure 6b shows the similarity in patterns of neural response within cluster (diagonal values) and between clusters (off-diagonal values). To determine whether each object cluster generated a distinct pattern of response, we compared the within-cluster (i.e., same condition) correlations with the between-cluster (i.e., different condition) correlations. This was performed separately for each background condition (textured, untextured). Distinct patterns of response to a cluster are demonstrated by higher within-cluster than between-cluster correlations.

We found that different object clusters evoked distinct patterns of neural response across the ventral visual ROI. A three-way analysis of variance (ANOVA), with background (untextured, textured), cluster (1–10) and comparison (within-cluster, between-cluster) as repeated measures showed a main effect of comparison ($F[9,180] = 173.59$,

$\eta_G^2 = 0.52$, $p < .001$), with within-cluster correlations being higher than between-cluster correlations. However, there was an interaction with background ($F[1,20] = 99.22$, $\eta_G^2 = 0.21$, $p < .001$), suggesting that the distinctiveness of cluster-specific patterns differed across background types (Figure 6c). Post hoc analysis revealed higher within-cluster than between-cluster correlations for both untextured ($t[20] = 13.95$, Cohen's $d = 3.04$, $p < .001$) and textured backgrounds ($t[20] = 8.88$, Cohen's $d = 1.94$, $p < .001$), but a stronger effect for untextured images ($t[20] = 10.25$, Cohen's $d = 2.24$, $p < .001$).

There was a two-way interaction between comparison and cluster ($F[9,180] = 3.80$, $\eta_G^2 = 0.04$, $p < .001$) and a three-way interaction between background, cluster, and comparison ($F[9,180] = 2.04$, $\eta_G^2 = 0.021$, $p = .037$). This shows that the distinctiveness of the neural patterns of response varies across clusters (Figure 6d). To investigate these effects, post hoc pairwise comparisons of the

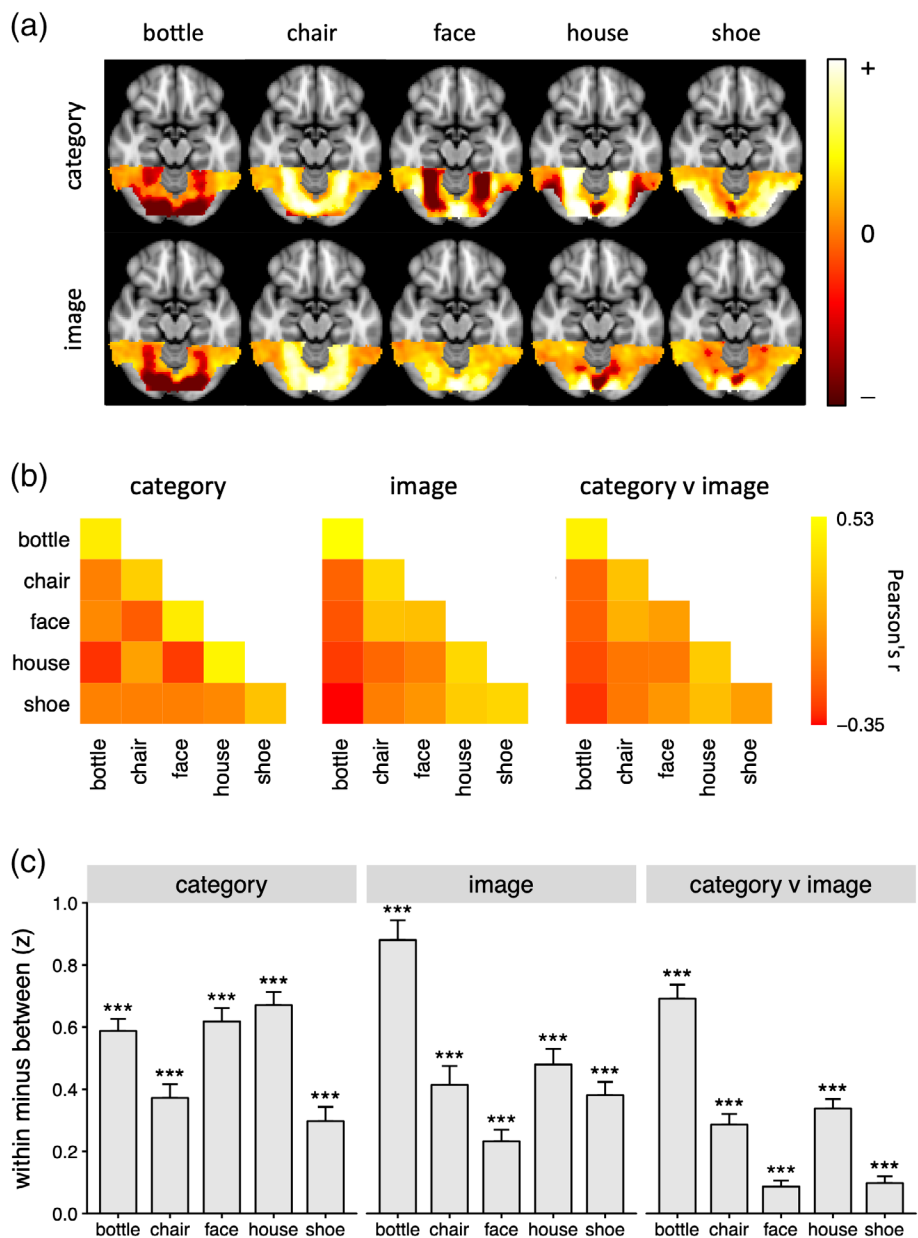


FIGURE 7 MVPA for Experiment 2. (a) Patterns of neural response in the ventral visual pathway to different object conditions. Patterns of response were normalised for each image type (row) by subtracting the voxel-wise mean response across five categories from the response to each category. Axial slices are located at $z = -16$ (ICBM-MNI 152). (b) Group mean neural matrices showing correlations between neural responses within and between the different object conditions. In the right matrix, correlations were performed across the 'category' and 'image' conditions. (c) Bar plot showing within-minus-between-category correlations (Z-transformed) for each image type. Error bars represent standard error of the mean. A similar pattern of results was evident when the data were analysed using a permutation test (Table S2). * $p < .05$, ** $p < .01$, *** $p < .001$ [Color figure can be viewed at wileyonlinelibrary.com]

within-cluster and between-cluster correlations were determined for each background-cluster combination. For the untextured background, there were significantly higher within-cluster than between-cluster correlations for all clusters ($t[20] > 6.5$, Cohen's $d > 1.53$, $p < .001$). For the textured background, nine of the 10 clusters showed higher within- than between-cluster correlations ($t[20] > 2.3$, Cohen's $d > 0.41$, $p < .022$ for 9 significant clusters. For the nonsignificant cluster (cluster 2), $t(20) = 1.75$, $p = .081$).

3.2 | Experiment 2

Experiment 2 compared the patterns of response to objects from five different categories (bottle, chair, face, house, and shoe) with the patterns of response to objects from other categories, but with similar image properties. Figure 7a and Figure S4 show the average pattern of response to each condition in Experiment 2. Figure 7b (left panel; category) shows the similarity in patterns of response within and between conditions to objects defined by category. Figure 7b (middle panel; image) shows the similarity in the patterns of response to metameric object clusters defined by image properties. Figure 7b (right panel; category vs. image) directly compares the similarity in the patterns of response to objects defined by category with the patterns of response to metameric object clusters defined by their image properties.

To determine whether the patterns of neural response to each object cluster were reliable, we compared the within-cluster correlations (on-diagonal values) with the between-cluster correlations (off-diagonal values). Distinct patterns of response to a cluster are demonstrated by higher within-cluster than between-cluster correlations. This effect of Comparison (within-cluster, between-cluster) was performed separately for each Image Type (category, image, and category/image) and for each Category (bottle, chair, face, house, and shoe).

A three-way analysis of variance (ANOVA), with Comparison, Image Type and Category as repeated measures showed a main effect of Comparison ($F[1,29] = 157.88$, $\eta_G^2 = 0.45$, $p < .001$), due to within-cluster correlations being higher than between-cluster correlations. Figure 7c (left panel) shows higher within-condition compared to between-condition for all comparisons between objects defined by category (bottle: $t[24] = 15.09$, Cohen's $d = 3.02$, $p < .001$; chair: $t[24] = 8.34$, Cohen's $d = 1.67$, $p < .001$; face: $t(24) = 14.35$, Cohen's $d = 2.87$, $p < .001$; house: $t(24) = 15.54$, Cohen's $d = 3.10$, $p < .001$; shoe: $t(24) = 6.55$, Cohen's $d = 1.31$, $p < .001$). Figure 7c (middle panel) also showed higher within-condition compared to between-condition comparisons for objects defined by image properties (bottle: $t(24) = 13.93$, Cohen's $d = 2.79$, $p < .001$; chair: $t(24) = 6.74$, Cohen's $d = 1.35$, $p < .001$; face: $t(24) = 6.39$, Cohen's $d = 1.28$, $p < .001$; house: $t(24) = 9.50$, Cohen's $d = 1.90$, $p < .001$; shoe: $t(24) = 8.87$, Cohen's $d = 1.77$, $p < .001$). Finally, Figure 7c (right panel) showed higher within-condition compared to between-condition comparisons for the cross-decoding of objects defined by category compared to objects defined by image properties (bottle: $t(24) = 15.41$, Cohen's

$d = 3.08$, $p < .001$; chair: $t(24) = 8.10$, Cohen's $d = 1.62$, $p < .001$; face: $t(24) = 4.52$, Cohen's $d = 0.90$, $p < .001$; house: $t(24) = 11.15$, Cohen's $d = 2.23$, $p < .001$; shoe: $t(24) = 4.45$, Cohen's $d = 0.89$, $p < .001$).

There was an interaction between Comparison, Image Type and Category ($F[1,20] = 91.20$, $\eta_G^2 = 0.17$, $p < .001$). This may reflect the lower correlations in the category versus image comparison (see Figure 7c), particularly the smaller effect of comparison for faces. One possible reason for this could be that image properties of the object clusters defined by image properties were similar, but not identical to the object clusters defined by category. To determine whether this could explain the variation in the neural similarity measures, we determined the mean correlation in image properties between objects defined by category and objects defined by image properties independently for each object category (see Figure 3). These five correlation values on image similarity were then compared to the similarity in the pattern of response across corresponding object clusters [e.g., bottle (category) vs. bottle (image)]. A regression analysis across participants showed that the similarity in image properties significantly predicted the neural similarity between the patterns of response between the category and image conditions ($t(24) = 9.07$, Cohen's $d = 1.81$, $p < .001$).

3.3 | Can image properties predict patterns of response in the ventral visual pathway?

Our next analysis investigated the extent to which the pattern of neural response in the ventral visual pathway could be predicted by the visual properties of the image clusters (Figure 8 and Figure S5-S6). This analysis was restricted to the between-condition values. In Experiment 1, there were significant, positive correlations between the neural correlation matrix and the image properties (untextured: $r = .43$, $p = .003$; textured: $r = .41$, $p = .005$), suggesting that clusters with more similar image properties were also likely to elicit more similar patterns of neural response (Figure 8a). Similarly, in Experiment 2, there was a significant positive correlation between image properties and patterns of neural response ($r = .74$, $p = .001$).

To determine the extent to which GIST descriptor explains all the variance in the neural data, we performed an additional hierarchical regression on the neural data shown in Figures 6b and 7b. Our aim was to establish if there was any residual within vs between condition variance, after removing the variance that could be explained by GIST. To do this, we first used GIST as a model, followed by a Within-Between model with values of 1 for within-condition comparisons and values of 0 for between-condition comparisons. In Experiment 1, there was a significant effect of the GIST model (untextured: mean $\beta = 0.397$, $t(20) = 14.0$, $p < .001$; textured: mean $\beta = 0.142$, $t(20) = 8.88$, $p < .001$). After accounting for GIST, there was a small but significant effect of the Within-Between model (untextured: mean $\beta = 0.034$, $t(20) = 6.79$, $p < .001$; textured: mean $\beta = 0.015$, $t(20) = 2.63$, $p = .016$). In Experiment 2, we ran a similar analysis on the category conditions. Again, there was a significant effect of the GIST model (mean $\beta = 1.239$, $t(20) = 19.17$, $p < .001$) and, after accounting for GIST there was a small, but significant effect of the

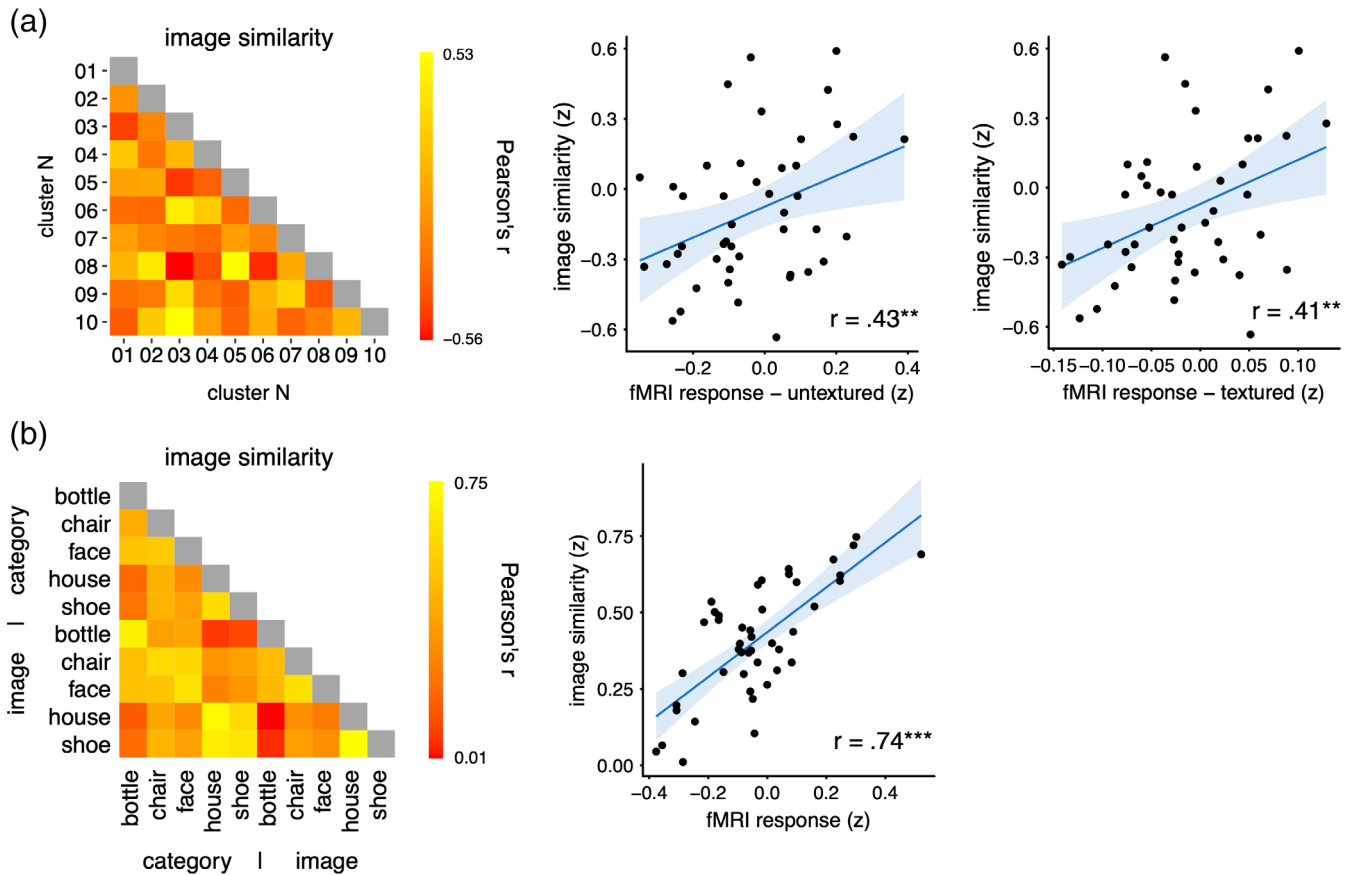


FIGURE 8 Representational similarity analysis between neural response and image properties for Experiment 1 (a) and Experiment 2 (b). Scatterplots show correlation between models and the neural matrices. Blue shaded region represents 95% confidence intervals. Prior to correlation, values in the image and neural matrices were Z-transformed and within-cluster correlations were removed. $*p < .05$, $**p < .01$, $***p < .001$ [Color figure can be viewed at wileyonlinelibrary.com]

Within-Between model (mean $\beta = 0.134$, $t(24) = 9.88$, $p < .001$). These analyses show that GIST does not predict all of the variance in the neural data. Given that this residual variance was evident in both Experiment 1 and 2, this suggests that the neural response is selective to image properties that are not fully captured by the GIST.

3.4 | The representation of objects across visual cortex

To explore how the neural representation of the object clusters changed along the visual processing hierarchy, we used probabilistic visual field map ROIs (Wang, Mruczek, Arcaro, & Kastner, 2015). These maps extend from early visual areas in the posterior occipital lobe to ventral and lateral regions of the temporal lobe. These ROIs allowed to us to perform a more fine-grained analysis of how image properties are represented across visual cortex, as compared to the ventral stream mask that may contain both mid- and high-level visual regions. We compared patterns of response to each image cluster to generate a similarity matrix for each region. This generated similarity matrices for the neural response to different object clusters across all the regions (cf. Figures 6b, 7b). For each region, the between-cluster similarity matrix was compared to each of the other regions. This

was done separately for Experiment 1 (Figure 7a) and Experiment 2 (Figure 7b) to create similarity matrices across regions. To determine how the regions were inter-connected a hierarchical clustering analysis was performed using an unweighted average distance method for computing the distance between clusters and 1-correlation value as the distance metric. This shows a division between the 'low-level' and 'high-level' visual regions, showing the emergence of a different neural representation of objects in 'high-level' regions. There was a significant correlation between the values across Experiment 1 and 2 ($r = .58$, $p < .001$).

4 | DISCUSSION

The aim of this study was to compare how objects are represented in the ventral visual pathway. A key feature of our approach to this question is the use of data-driven methods for image selection. During a life-time of natural viewing, a person encounters a vast number of objects. However, during a typical neuroimaging experiment, only a finite number of images can be presented. Thus, the stimuli selected may not sample image space in a uniform and objective way, making it difficult to separate the effects of arbitrary and subjective

manipulations of stimulus conditions from those driven by the fundamental underlying dimensions. To address this issue, we used a data-driven approach in which a clustering algorithm was used to evenly sample clusters of objects from a large database of images. Images were defined based on their image properties to avoid the need to impose any additional higher-level constraints. Our rationale is that these object clusters will provide a good first approximation to the diversity of objects that an individual has been exposed to during a lifetime of natural viewing.

We found that object clusters with similar image properties gave rise to distinct patterns of neural response in the ventral visual pathway. This is consistent with previous studies showing that image properties predict patterns of response to objects in the ventral visual pathway (Andrews et al., 2015; Rice et al., 2014; Watson et al., 2016). In these previous studies, the image conditions were from the same category, so it is possible that the similarity in image properties could have been confounded with correlated differences in semantic properties. In this study, the images in each cluster did not have any consistent semantic properties, which reinforce the importance of image properties in the neural representation of this region. Another important aspect of the analysis was that the similarity in the patterns of response in the ventral visual pathway could be predicted by the similarity in their image properties. This strong linkage shown in Figure 8 provides a strong test of the hypothesis that an image-based representation of objects is evident in this region.

Cluster-specific patterns of neural response in the ventral visual pathway were less distinct when images were imposed on a textured background, relative to an untextured background (see Figure 6c). An important difference between these two conditions is the contrast-defined spatial envelope or outline of the object. In the untextured condition, this is identical to the spatial boundary of the object, which differs systematically across object clusters. However, all objects in the textured condition were presented within a square of pink noise, reducing the salience of this diagnostic cue. The textured backgrounds were designed to provide visual stimulation across the extent of the image and emulate the $1/f$ amplitude energy distribution found in most natural images. The reduction in the distinctiveness of cluster-specific responses when a textured background is added suggests that the spatial envelope is an important visual feature in determining the topographic response of the ventral visual pathway (Bracci & Op de Beeck, 2016; Vernon, Gouws, Lawrence, Wade, & Morland, 2016; Watson et al., 2016). Although the use of these textured backgrounds avoids conveying any confounding information, we expect that presenting objects on structured backgrounds that are randomly selected would have given similar results, due to suppression of the spatial envelope (Yamins et al., 2014). Nevertheless, the similarity matrix for objects on untextured and textured backgrounds was highly correlated. This, along with the persistence of attenuated, but distinctive, cluster responses in the presence of a textured background suggests that the neural patterns to untextured images generalise to natural images in which the ability to separate figure and ground is likely to be an important processing step (Rubin, 2001).

As a further test of an image-based representation, we asked whether category-selective patterns of response in the ventral visual

pathway could be generated by objects from different categories, but with similar image properties. In other words, is the pattern of response to bottles similar to the pattern of response to objects that have similar image properties to bottles. To do this, we measured the image properties from exemplars of different categories (bottles, chairs, faces, houses, and shoes) and then selected objects with similar image properties (excluding any objects from the original categories). Although the objects in each metameric cluster did not have any consistent semantic properties, we found that they elicited distinct patterns of response in the ventral visual pathway. Moreover, the pattern of response to each metameric object cluster was similar to the pattern of response elicited by the corresponding category-defined, object cluster. For example, the pattern of response to chairs was similar to the pattern of response to objects that had similar image properties to chairs. We found that the ability to find objects that had similar image properties to a category, but were not a member of that category varied for different categories. For some natural categories, such as faces, visual properties are more distinctive and consistent than others, meaning that degree of similarity between the matched images varied across conditions. Interestingly, this variation predicted similarity in the pattern of neural response; images with more similar properties generated more similar patterns of neural response.

The fact that low-level properties of objects can predict patterns of response in 'high-level' regions does not imply that information is represented in a similar way to 'low-level' or early visual areas. In fact, our data clearly shows that the neural representation changes along the visual hierarchy (see Figure 9). An important property of natural images is that they contain strong statistical dependencies, such as location-specific combinations of orientation and spatial frequency corresponding to image features such as edges. Indeed, the character and extent of these statistical dependencies are likely to be diagnostic for different classes of objects (Geisler, 2008; Oppenheim & Lim, 1981; Sigman, Cecchi, Gilbert, & Magnasco, 2001; Thomson, 1999). Although we found that GIST was able to predict the majority of the variance in the patterns of neural response in the ventral visual pathway, a hierarchical regression analysis found that there was some within-condition variance in the neural patterns that was not explained. Given this, it is possible that other models that incorporate mid-level representations of objects may predict patterns of neural response more accurately than GIST (Guclu & van Gerven, 2015; Khaligh-Razavi & Kriegeskorte, 2014; Leeds, Seibert, Pyles, & Tarr, 2013; Long et al., 2018; Yamins et al., 2014). This is consistent with models of object processing in which selectivity for objects emerges through the superposition of topographically organised maps representing lower-level properties (Andrews et al., 2015; Op de Beeck, Haushofer, & Kanwisher, 2008). The response to particular object categories may reflect the convergence of selectivity for particular combinations of image properties that are diagnostic of that object category. The image properties of objects that are more commonly seen may be over-represented in high-level regions in the same way that the central visual field is over-represented in low-level visual regions.

An obvious advantage of a relatively image-based representation in high-level visual cortex is that it can be used more flexibly in the

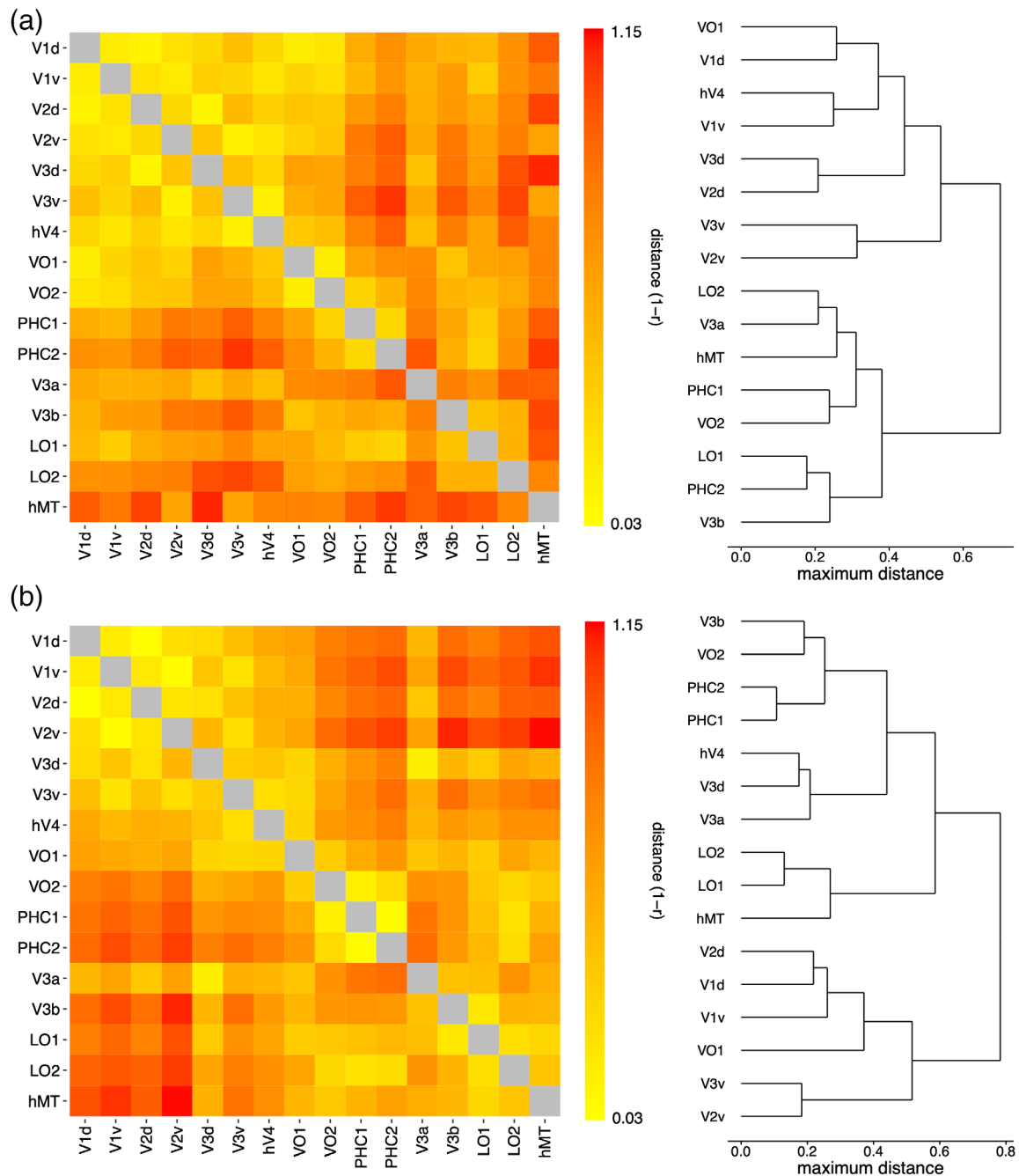


FIGURE 9 Comparison of retinotopic regions for Experiment 1 (a) and Experiment 2 (b). Matrices show representational similarity of retinotopic regions based on the neural similarity matrices. Dendrograms show hierarchical clustering of regions based on maximum distance [Color figure can be viewed at wileyonlinelibrary.com]

processing of objects. Previous studies have shown that patterns of neural response in the ventral visual pathway can discriminate higher-level properties of objects (Grill-Spector & Weiner, 2014), such as category (Connolly et al., 2012; Haxby et al., 2001; Kriegeskorte et al., 2008; Naselaris et al., 2009), animacy (Chao, Haxby, & Martin, 1999; Kriegeskorte et al., 2008) and real-world size (Konkle & Oliva, 2012). Our results suggest that these higher-level representations are linked to correlated variation in low-level properties of objects. This implies that the ventral visual pathway could have a fundamentally image-

based representation, albeit biased toward those features that are critical for perception. A more fine-grained analysis of how the neural representation of image properties changes from low- to mid- to high-level regions is shown in Figure 9. This analysis shows that image properties are represented differently in different regions along the visual hierarchy. This presumably reflects differences in the complexity of the image based representation. So, rather than a transition from low-level to high-level properties, there is a gradual change in the complexity by which image properties are represented (see

Coggan et al., 2017). This may covary with higher-level dimensions, particularly in more anterior regions. However, a fundamentally image-based representation would allow for the extraction of different information depending on the task.

An important feature of our findings is that the spatial patterns of response to different object clusters generalised across participants. Neuroimaging studies have shown that the locations of category-selective regions in the ventral visual pathway are broadly consistent across individuals (Kanwisher, 2010). This implies that common principles may well underpin the organisation of these regions. In many neuroimaging studies, MVPA is performed at the individual participant level. This approach is often grounded in an assumption of substantial differences between individual brains, and contrasts with the across-participant analysis used in the current study. In our analysis, we compared the pattern of response in individual participants with the pattern from a group analysis in which that participant was left out (Coggan et al., 2017; Flack et al., 2015; Rice et al., 2014; Watson et al., 2014; Weibert et al., 2018). The success of this approach shows that much of the topographic pattern of response to natural images is consistent across individuals. These observations are significant in that they suggest that our findings reflect the operation of large-scale organising principles that are consistent across different individuals.

In summary, we used a data-driven approach to group images of objects into different clusters based on their visual properties. This circumvents the limitations associated with subjectively allocating stimuli to predefined categories. Although the clusters did not correspond to typical object categories, we found that they elicited distinct patterns of response in the ventral visual pathway. The results also show how category-selective patterns of response can be explained by the image properties of objects. Interestingly, the representational structure found in 'high-level' regions was not the same as that found in 'low-level' regions. This suggests the emergence of an image-based representation in high-level visual cortex that is based on the statistical properties of objects. Although we have used image properties to select images, it would also be possible to extend this approach to the selection of objects based on other low-, mid-, or high-level properties.

CONFLICT OF INTERESTS

The authors declare no competing financial interests.

DATA ACCESSIBILITY

The R code and representational similarity matrices are available at <https://github.com/ddcoggan/p007>.

ORCID

David D. Coggan  <https://orcid.org/0000-0001-9090-7084>

Daniel H. Baker  <https://orcid.org/0000-0002-0161-443X>

Timothy J. Andrews  <https://orcid.org/0000-0001-8255-9120>

REFERENCES

- Andrews, T. J., Clarke, A., Pell, P., & Hartley, T. (2010). Selectivity for low-level features of objects in the human ventral stream. *NeuroImage*, 49(1), 703–711.
- Andrews, T. J., Watson, D. M., Rice, G. E., & Hartley, T. (2015). Low-level properties of natural images predict topographic patterns of neural response in the ventral visual pathway. *Journal of Vision*, 15(7), 3.
- Bellman, R. E. (1961). *Adaptive control processes: A guided tour*. London: Princeton University Press.
- Bonhoeffer, T., & Grinvald, A. (1991). Iso-orientation domains in cat visual cortex are arranged in pinwheel-like patterns. *Nature*, 353(6343), 429–431.
- Bracci, S., & Op de Beeck, H. (2016). Dissociations and associations between shape and category representations in the two visual pathways. *Journal of Neuroscience*, 36(2), 432–444.
- Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The bank of standardized stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PLoS ONE*, 5(5), e10773.
- Chao, L., Haxby, J. V., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neuroscience*, 2(10), 913–919.
- Clarke, A., & Tyler, L. (2014). Object-specific semantic coding in human Perirhinal cortex. *The Journal of Neuroscience*, 34(14), 4766–4775.
- Coggan, D. D., Allen, L. A., Farrar, O. R. H., Gouws, A. D., Morland, A. B., Baker, D. H., & Andrews, T. J. (2017). Differences in selectivity to natural images in early visual areas (V1 - V3). *Scientific Reports*, 7(2444), 1–8.
- Coggan, D. D., Baker, D. H., & Andrews, T. J. (2016). The role of visual and semantic properties in the emergence of category-specific patterns of neural response in the human brain. *ENeuro*, 3(4), ENEURO.0158-16.2016. <https://doi.org/10.1523/ENEURO.0158-16.2016>.
- Coggan, D. D., Liu, W., Baker, D. H., & Andrews, T. J. (2016). Category-selective patterns of neural response in the ventral visual pathway in the absence of categorical information. *NeuroImage*, 135, 107–114.
- Coggan, D. D., Baker, D. H., & Andrews, T. J. (2019). Selectivity for mid-level properties of faces and places in the fusiform face area and parahippocampal place area. *European Journal Neuroscience*, 49, 1587–1596. <https://doi.org/10.1111/ejn.14327>
- Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M. A., & Michel, F. (2000). The visual word form area: Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain: A Journal of Neurology*, 123(2), 291–307.
- Connolly, A. C., Guntupalli, J. S., Gors, J., Hanke, M., Halchenko, Y. O., Wu, Y.-C., ... Haxby, J. V. (2012). The representation of biological classes in the human brain. *Journal of Neuroscience*, 32(8), 2608–2618.
- Downing, P. E., Chan, A. W.-Y., Peelen, M., Dodds, C. M., & Kanwisher, N. (2006). Domain specificity in visual cortex. *Cerebral Cortex*, 16(October 2006), 1453–1461.
- Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science (New York, N.Y.)*, 293(5539), 2470–2473.
- Engel, S. A., Wandell, B. A., Rumelhart, D. E., Lee, A. T., Glover, G. H., Chichilnisky, E. J., & Shadlen, M. N. (1994). fMRI of human visual cortex. *Nature*, 369, 525.
- Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, 392(6676), 598–601.
- Flack, T. R., Andrews, T. J., Hymers, M., Al-Mosaiwi, M., Marsden, S. P., Strachan, J. W. A., ... Young, A. W. (2015). Responses in the right posterior superior temporal sulcus show a feature-based response to facial expression. *Cortex*, 69, 14–23.
- Geisler, W. S. (2008). Visual perception and the statistical properties of natural scenes. *Annual Review of Psychology*, 59, 167–192.

- Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews. Neuroscience*, 15(8), 536–548.
- Guclu, U., & van Gerven, M. A. J. (2015). Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *The Journal of Neuroscience*, 35, 10005–100014.
- Hanke, M., Halchenko, Y. O., Sederberg, P. B., Hanson, S. J., Haxby, J. V., & Pollmann, S. (2009). PyMVPA: A python toolbox for multivariate pattern analysis of fMRI data. *Neuroinformatics*, 7(1), 37–53.
- Hanson, S. J., Matsuka, T., & Haxby, J. V. (2004). Combinatorial codes in ventral temporal lobe for object recognition: Haxby (2001) revisited: Is there a “face” area? *NeuroImage*, 23(1), 156–166.
- Haxby, J. V., Gobbini, M., Furey, M., Ishai, A., Schouten, J., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science (New York, N.Y.)*, 293(5539), 2425–2430.
- Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, 195, 215–243.
- Jenkins, R., White, D., Van Montfort, X., & Burton, M. (2011). Variability in photos of the same face. *Cognition*, 121(3), 313–323.
- Jozwik, K. M., Kriegeskorte, N., & Mur, M. (2016). Visual features as stepping stones towards semantics: Explaining object similarity in IT and perception with non-negative least squares. *Neuropsychologia*, 83, 201–226.
- Kanwisher, N. (2010). Functional specificity in the human brain: A window into the functional architecture of the mind. *Proceedings of the National Academy of Sciences of the United States of America*, 107(25), 11163–11170.
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience*, 17(11), 4302–4311.
- Khaligh-Razavi, S. M., & Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS Computational Biology*, 10, e1003915.
- Konkle, T., & Oliva, A. (2012). A real-world size Organization of Object Responses in Occipitotemporal cortex. *Neuron*, 74(6), 1114–1124.
- Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., ... Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, 60(6), 1126–1141.
- Leeds, D. D., Seibert, D. A., Pyles, J. A., & Tarr, M. J. (2013). Comparing representations across human fMRI and computational vision. *Journal of Vision*, 13(13), 25.
- Lescroart, M. D., Gallant, J. L. (2019). Human Scene-Selective Areas Represent 3D Configurations of Surfaces. *Neuron*, 101(1), 178–192.e7. <https://doi.org/10.1016/j.neuron.2018.11.004>.
- Levy, I., Hasson, U., Avidan, G., Hendler, T., & Malach, R. (2001). Center-periphery organization of human object areas. *Nature Neuroscience*, 4, 533–539.
- Long, B., Yu, C.-P., & Konkle, T. (2018). Mid-level visual features underlie the high-level categorical organization of the ventral stream. *Proceedings of the National Academy of Sciences*, 115(38), E9015–E9024.
- Malcolm, G. L., Groen, I. I. A., Baker, C. I. (2016). Making Sense of Real-World Scenes. *Trends in Cognitive Sciences*, 20(11), 843–856. <https://doi.org/10.1016/j.tics.2016.09.003>.
- McCarthy, G., Puce, A., Gore, J. C., & Truett, A. (1997). Face-specific processing in the human fusiform Gyrus. *Journal of Cognitive Neuroscience*, 9(5), 605–610.
- McNeil, J. E., & Warrington, E. K. (1993). Prosopagnosia: A face-specific disorder. *The Quarterly Journal of Experimental Psychology Section A*, 46(1), 1–10.
- Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action*. Oxford: Oxford University Press.
- Moscovitch, M., Winocur, G., & Behrmann, M. (1997). What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. *Journal of Cognitive Neuroscience*, 9(5), 555–604.
- Nasr, S., Echavarria, C. E., & Tootell, R. B. H. (2014). Thinking Outside the Box: Rectilinear Shapes Selectively Activate Scene-Selective Cortex. *Journal of Neuroscience*, 34(20), 6721–6735. <https://doi.org/10.1523/JNEUROSCI.4802-13.2014>
- Naselaris, T., Prenger, R. J., Kay, K. N., Oliver, M., & Gallant, J. L. (2009). Bayesian reconstruction of natural images from human brain activity. *Neuron*, 63(6), 902–915.
- Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42(3), 145–175.
- Op de Beeck, H. P., Haushofer, J., & Kanwisher, N. G. (2008). Interpreting fMRI data: Maps, modules and dimensions. *Nature Reviews. Neuroscience*, 9(2), 123–135.
- Oppenheim, A. V., & Lim, J. S. (1981). “The importance of phase in signals,” in Proceedings of the IEEE. 69(5), 529–541. <https://doi.org/10.1109/PROC.1981.12022>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, É. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Peirce, J. W. (2007). PsychoPy-psychophysics software in python. *Journal of Neuroscience Methods*, 162(1), 8–13.
- Poldrack, R. A., Halchenko, Y. O., & Hanson, S. J. (2009). Decoding the large-scale structure of brain function by classifying mental states across individuals. *Psychological Science*, 20(11), 1364–1372.
- Proklova, D., Kaiser, D., & Peelen, M. (2016). Disentangling representations of object shape and object category in human visual cortex: The animate–inanimate distinction. *Journal of Cognitive Neuroscience*, 38(5), 680–692.
- Rice, G. E., Watson, D. M., Hartley, T., & Andrews, T. J. (2014). Low-level image properties of visual objects predict patterns of neural response across category-selective regions of the ventral visual pathway. *The Journal of Neuroscience*, 34(26), 8837–8844.
- Rubin, N. (2001). Figure and ground in the brain. *Nature Neuroscience*, 4(9), 857–858.
- Sigman, M., Cecchi, G. A., Gilbert, C. D., & Magnasco, M. O. (2001). On a common circle: Natural scenes and Gestalt rules. *Proceedings of the National Academy of Sciences*, 98(4), 1935–1940. <https://doi.org/10.1073/pnas.98.4.1935>
- Sormaz, M., Watson, D. M., Smith, W. A. P., Young, A. W., & Andrews, T. J. (2016). Modelling the perceptual similarity of facial expressions from image statistics and neural responses. *NeuroImage*, 129, 64–71.
- Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience*, 19, 109–139.
- Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behaviour* (pp. 549–586). Cambridge, MA: MIT Press.
- Vernon, R. J. W., Gouws, A. D., Lawrence, S. J. D., Wade, A. R., & Morland, A. B. (2016). Multivariate patterns in the human object-processing pathway reveal a shift from Retinotopic to shape curvature representations in lateral occipital areas, LO-1 and LO-2. *Journal of Neuroscience*, 36(21), 5763–5774.
- Vul, E., Lashkari, D., Hsieh, P.-J., Golland, P., & Kanwisher, N. (2012). Data-driven functional clustering reveals dominance of face, place, and body selectivity in the ventral visual pathway. *Journal of Neurophysiology*, 108(8), 2306–2322.
- Wandell, B. A., & Winawer, J. (2011). Imaging retinotopic maps in the human brain. *Vision Research*, 51, 718–737.
- Wang, L., Mruczek, R. E. B., Arcaro, M. J., & Kastner, S. (2015). Probabilistic maps of visual topography in human cortex. *Cerebral Cortex*, 25(10), 3911–3931.
- Watson, D. M., Andrews, T. J., & Hartley, T. (2017). A data driven approach to understanding the organization of high-level visual cortex. *Scientific Reports*, 7(1), 3596.

- Watson, D. M., Hartley, T., & Andrews, T. J. (2014). Patterns of response to visual scenes are linked to the low-level properties of the image. *NeuroImage*, *99*, 402–410.
- Watson, D. M., Young, A. W., & Andrews, T. J. (2016). Spatial properties of objects predict patterns of neural response in the ventral visual pathway. *NeuroImage*, *126*(2016), 173–183.
- Weibert, K., Flack, T. R., Young, A. W., & Andrews, T. J. (2018). Patterns of neural response in face regions are predicted by low-level image properties. *Cortex*, *103*, 199–210.
- Yamins, D. L. K., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, 8619–8624.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Coggan DD, Giannakopoulou A, Ali S, et al. A data-driven approach to stimulus selection reveals an image-based representation of objects in high-level visual areas. *Hum Brain Mapp.* 2019;1–16. <https://doi.org/10.1002/hbm.24732>