NeuroImage 135 (2016) 107–114

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Category-selective patterns of neural response in the ventral visual pathway in the absence of categorical information

David D. Coggan, Wanling Liu, Daniel H. Baker, Timothy J. Andrews *

Department of Psychology and York Neuroimaging Centre, University of York, York YO10 5DD, United Kingdom

ARTICLE INFO

Article history:
Received 22 August 2015
Revised 21 April 2016
Accepted 26 April 2016
Available online 28 April 2016

Keywords:
Face
fMRI
MVPA
Object recognition
Ventral stream

ABSTRACT

Neuroimaging studies have revealed distinct patterns of response to different object categories in the ventral visual pathway. These findings imply that object category is an important organizing principle in this region of visual cortex. However, object categories also differ systematically in their image properties. So, it is possible that these patterns of neural response could reflect differences in image properties rather than object category. To differentiate between these alternative explanations, we used images of objects that had been phase-scrambled at a local or global level. Both scrambling processes preserved many of the lower-level image properties, but rendered the images unrecognizable. We then measured the effect of image scrambling on the patterns of neural response within the ventral pathway. We found that intact and scrambled images evoked distinct category-selective patterns of activity in the ventral stream. Moreover, intact and scrambled images of the same object category produced highly similar patterns of response. These results suggest that the neural representation in the ventral visual pathway is tightly linked to the statistical properties of the image.

1. Introduction

Visual areas involved in object perception form a ventral processing pathway that projects from the occipital toward the temporal lobe (Ungerleider and Mishkin, 1982; Milner and Goodale, 1995). Lesions to these regions of the brain often result in difficulties in the perception and recognition of different categories of objects (McNeil and Warrington, 1993; Moscovitch et al., 1997). Consistent with these neuropsychological reports, neuroimaging studies have shown that discrete regions of the ventral visual pathway are specialized for different categories of objects. For example, while some regions of the ventral visual pathway are more responsive to images of faces than to images of non-face objects (Kanwisher et al., 1997; McCarthy et al., 1997), other regions are selective for images of places (Epstein and Kanwisher, 1998), body parts (Downing et al., 2001) and visually presented words (Cohen et al., 2000). This selectivity has been regarded as characteristic of a modular organization in which distinct regions process specific object categories (Kanwisher, 2010).

Despite the evidence for category-selectivity in the ventral visual pathway, specialized regions have only been reported for a limited number of object categories (Downing et al., 2006; Op de Beeck et al., 2008). However, other studies using multivariate fMRI analysis methods have demonstrated that the spatial pattern of response across the entire ventral stream can distinguish a greater range of object categories (Ishai et al., 1999; Haxby et al., 2001; Kriegeskorte et al., 2008). The distributed nature of the fMRI response to different categories of objects within the ventral visual pathway has been interpreted as showing a topographic map of object form (Haxby et al., 2001).

The topographic organization of the ventral stream is thought to be analogous with the topographic organization found in early visual areas, in which responses are tightly linked to low-level properties of the image, such as spatial frequency, orientation and spatial position (Hubel and Wiesel, 1968; Bonhoeffer and Grinvald, 1991; Engel et al., 1994; Wandell et al., 2007). In contrast to early visual areas, a variety of evidence has suggested that patterns of response in the ventral visual pathway are linked to the categorical or semantic information that the images convey rather than to the image properties (Kriegeskorte et al., 2008; Naselaris et al., 2009; Connolly et al., 2012). Evidence for other organizing principles can be found in the large-scale patterns of response to animacy (Chao et al., 1999; Kriegeskorte et al., 2008) and the real-world size of objects (Konkle and Oliva, 2012). However, it remains unclear how the selectivity for high-level properties in ventral visual pathway might arise from image-based representations found in early visual regions (Op de Beeck et al., 2008). In recent studies, we showed that low-level properties of objects could predict patterns of neural response in the ventral visual pathway (Rice et al., 2014; Watson et al., 2016). However, images drawn from the same object category are likely to have similar low-level properties. So, the link between image properties and patterns of neural response is expected under both categorical and image-based accounts.

The aim of the present study was to directly determine whether the category-selective patterns of neural response across the ventral visual pathway can be explained by selectivity to more basic properties of

* Corresponding author.
E-mail address: timothy.andrews@york.ac.uk (T.J. Andrews).

http://dx.doi.org/10.1016/j.neuroimage.2016.04.060
1053-8119/

the stimulus. To address this question, we measured the neural response across the ventral visual pathway to intact images of different object categories, as well as versions of these images that had been phase-scrambled on a global or local basis. Our rationale for using scrambled images is that they have many of the image properties found in intact images, but do not convey any categorical or semantic information, thus providing dissociation between higher-level and lower-level information. Our hypothesis was that, if neurons in the ventral stream are selective for the categorical or semantic properties conveyed by the image, there should be no correspondence between patterns of response evoked by intact and scrambled images. Conversely, if patterns of response in the ventral stream reflect selectivity to more basic dimensions of the stimulus, we would predict a significant correlation between patterns of response to intact and scrambled images.

2. Materials and methods

2.1. Stimuli

180 images of five object categories (bottles, chairs, faces, houses, shoes) were taken from an object image stimulus set (Rice et al., 2014). All images were grey-scale, superimposed on a mid-grey background, and had a resolution of 400 × 400 pixels. Images were viewed at a distance of 57 cm and subtended 8° of visual angle. For each original image, two different phase-scrambled versions were generated. A global-scrambling method involved a typical Fourier-scramble, i.e. keeping the global power of each two-dimensional frequency component constant while randomizing the phase of the components. A local-scrambling method involved windowing the original image into an 8 × 8 grid in image space and applying a phase-scramble to each window independently. Examples of the images are shown in Fig. 1.

2.2. Behavioural study

In order to measure the effect of scrambling on the categorical information conveyed by an image, a behavioural study (approved by the Ethics Committee of the Psychology Department, University of York) involving an object naming task was conducted. Twenty-one participants took part (8 male, mean age = 28.4, SD = 14.5 years), none of whom participated in the fMRI study. All observers had normal or corrected-to-normal vision and gave written, informed consent. Stimuli were presented in 3 blocks. The first block contained globally scrambled images, the second block contained locally scrambled images and the third block contained intact images. Therefore, participants were unaware of the object categories in our stimulus set prior to viewing the scrambled images. Each block contained 25 trials. On each trial, participants fixated a cross in the centre of the screen for 200 ms before the stimulus was presented for 800 ms. If the participant failed to perceive an object in the image, they were instructed to move on to the next trial with a key-press. If the participant perceived an object in the image, they were instructed to write down the name of the object on an answer sheet and state their confidence in the accuracy of their answer. The confidence measure involved a 5-point scale ranging from 0 (extremely unsure) to 4 (certain). Correct responses were given to any answers that indicated that the observer had abstracted any accurate semantic or categorical information associated with the object. This ensured that accuracy scores represented an upper estimate of their ability to recognize the images. Analyses on accuracy and confidence data involved comparison of 95% confidence intervals (CI) using bootstrapping.

2.3. fMRI study

Twenty-two participants took part in the fMRI experiment (7 male, mean age = 23.0, SD = 1.4 years). Sample size was based on a previous study using a similar design (Rice et al., 2014). Data were collected from all participants prior to analysis. 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. Images were viewed on a screen at the rear of the scanner via a mirror placed immediately above the participant's head.

The fMRI experiment consisted of three scans, each lasting 7.5 min. The first scan contained globally scrambled images, the second scan contained locally scrambled images and the third scan contained intact images. In all scans, object categories were presented in blocks. There were 6 repetitions of each category in each scan. In each block, 6 images from a category were presented individually for 800 ms, with a 200 ms inter-stimulus-interval. This was followed by a fixation cross lasting 9 s. Participants performed a task while viewing images, designed to maintain attention for the duration of the scan, and be of equivalent difficulty across category and image type. The task consisted of pressing a button on a response box whenever a red dot appeared on an image, which occurred on either the 3rd, 4th, 5th or 6th image in each block.

All fMRI data were acquired with a General Electric 3T 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.

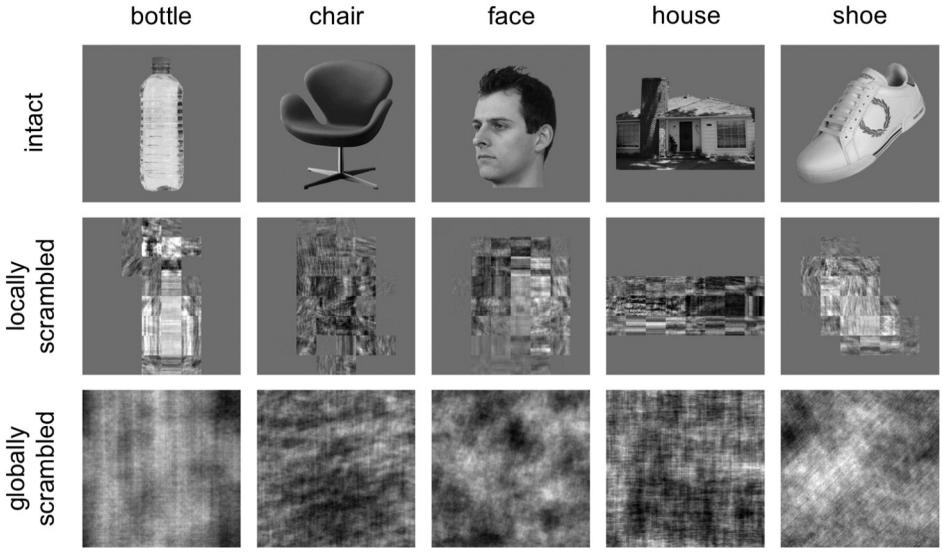

Fig. 1. Exemplars of intact, locally scrambled and globally scrambled images from the different object categories.

A gradient-echo echo-planar imaging (EPI) sequence was used to collect data from 38 contiguous axial slices (TR = 3000 ms, TE = 32.7 ms, FOV = 288 × 288 mm, matrix size = 128 × 128, slice thickness = 3 mm). The fMRI data were initially analysed with FEAT v5.98 (http://www.fmrib.ox.ac.uk/fsl). In all scans the initial 9 s of data was 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 condition by regressing the hemodynamic response of each voxel against a boxcar function convolved with a single-gamma HRF. Next, individual participant data were entered into higher-level group analyses using a mixed-effects design (FLAME, FSL). Functional data were first registered to a high-resolution T1-anatomical image and then onto the standard MNI brain (ICBM152).

To construct a mask of the ventral visual pathway, we selected a series of anatomical regions of interest (ROIs) from the Harvard-Oxford cortical atlas based on the physical limits of ventral temporal cortex described by Grill-Spector and Weiner (2014). Specifically, these regions were: inferior temporal gyrus (temporo-occipital portion), temporal–occipital fusiform cortex, occipital fusiform gyrus, and lingual gyrus. The overall ventral temporal mask was defined by a concatenation of the individual anatomical masks (see Fig. 3 inset).

Next, we measured patterns of response to the different stimulus conditions. For each participant, parameter estimates were generated for each category in each scan. The reliability of response patterns was tested using a leave-one-participant-out (LOPO) cross-validation paradigm (Poldrack et al., 2009; Rice et al., 2014; Watson et al., 2014) in which, for each individual parameter estimate, a corresponding group parameter estimate was determined using a higher-level analysis of the remaining participants. Parameter estimates were normalized by subtracting the mean response per voxel across all categories. This was performed separately for each scan. These data were then submitted to a correlation-based multi-variate pattern analyses (MVPA, Haxby et al., 2001, 2014) implemented using the PyMVPA toolbox (http://www.pymvpa.org/; Hanke et al., 2009). For each iteration of the LOPO cross-validation, the normalized patterns of response to each stimulus condition were correlated between the left-out participant and the remaining group. This allowed us to determine whether there were reliable patterns of response that were consistent across individual participants. A Fisher's Z-transformation was then applied to the correlations prior to further statistical analyses, and a Holm–Bonferroni correction was used to control the family-wise error rate across post hoc pairwise comparisons.

3. Results

First, we conducted a behavioural experiment to determine how image scrambling affected the categorical and semantic information that the images conveyed (Fig. 2). Mean accuracy for globally scrambled (mean = 0.8%, CI: 0.0–2.1%) and locally scrambled (mean = 6.9%, CI: 3.4–10.9%) images was significantly lower than for the intact (mean = 98.4%, CI: 96.6–99.6%) images. Analysis of the confidence ratings for correct answers showed that participants were significantly more confident in their responses to intact images (mean = 3.93, CI: 3.83–3.99) compared to locally scrambled (mean = 0.98, CI: 0.59–1.35) and globally scrambled (mean = 0.83, CI: 0.67–1.00) images. There was no significant difference in confidence ratings between globally scrambled and locally scrambled images. Finally, we contrasted the confidence ratings for correct and incorrect responses to locally scrambled images. There was no significant difference in confidence between correct (mean = 0.97, CI: 0.57–1.35) and incorrect (mean = 0.97, CI: 0.57–1.42) responses. This shows that both scrambling processes substantially reduce the categorical information conveyed by an image. It also suggests that, for correct responses to locally scrambled images, participants showed no more confidence than when they were incorrect or when they were viewing globally scrambled images.

Next, we measured patterns of ventral response to intact, locally scrambled and globally scrambled images from different object categories. Fig. 3 shows the normalized group responses to each condition across the ventral visual pathway. Responses above the mean are shown in red/yellow, and responses below the mean are shown in blue/light-blue. This shows that there are clear differences in the patterns of response to different categories of objects. However, the data also appear to show that the patterns of response are more similar between intact and scrambled images from the same category.

To quantify the reliability of these patterns across participants, we used correlation-based MVPA. Fig. 4A shows the group correlation matrices for each image type. To determine the reliability of the patterns, correlation values across all individual correlation matrices were entered into a repeated-measures ANOVA with Comparison (within-category, between-category), Image Type (intact, locally scrambled, globally scrambled) and Category (bottle, chair, face, house, shoe) as the main factors. There was a main effect of Comparison ($F(1, 21) = 354.50$, $p < .0001$). This was due to higher within-category correlations (e.g. bottle–bottle) compared to between-category correlations (e.g. bottle–chair). There were also main effects of Image Type ($F(2, 42) = 83.74$, $p < .0001$) and Category ($F(4, 84) = 6.26$, $p = .0002$), and a significant interaction between Image Type, Comparison and Category ($F(8, 168) = 6.12$, $p < .0001$). This interaction suggests that the distinctiveness of response patterns differs across image types and categories. To explore this, we analysed the data independently for each Image Type (see Fig. 4B) using an ANOVA with Comparison (within-category, between-category) and Category (bottle, chair, face, house, shoe) as the main factors.

For intact images, there was a main effect of Comparison ($F(1, 21) = 259.91$, $p < .0001$), due to higher within-category compared to between-category correlations. There was also an interaction between Comparison and Category ($F(4, 84) = 14.08$, $p < .0001$). Pairwise comparisons revealed significantly higher within-category compared to between-category correlations for bottles ($t(21) = 11.62$, $p < .0001$),

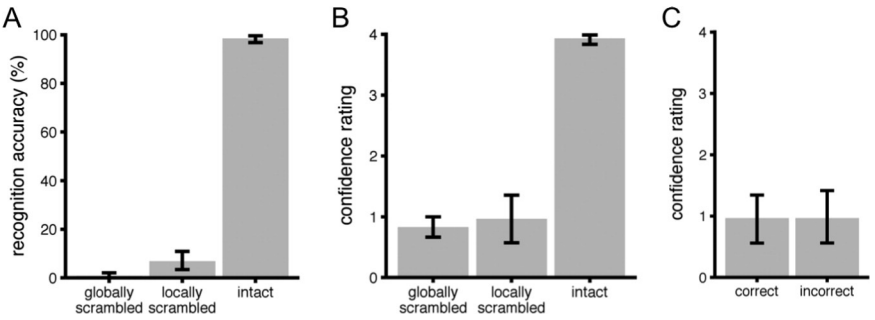

Fig. 2. Results of semantic naming task. (A) Mean accuracy scores for each image type. (B) Mean confidence ratings for correct responses for each image type. (C) Mean confidence ratings for correct and incorrect responses to locally scrambled images. For all panels, error bars reflect bootstrapped 95% confidence intervals.

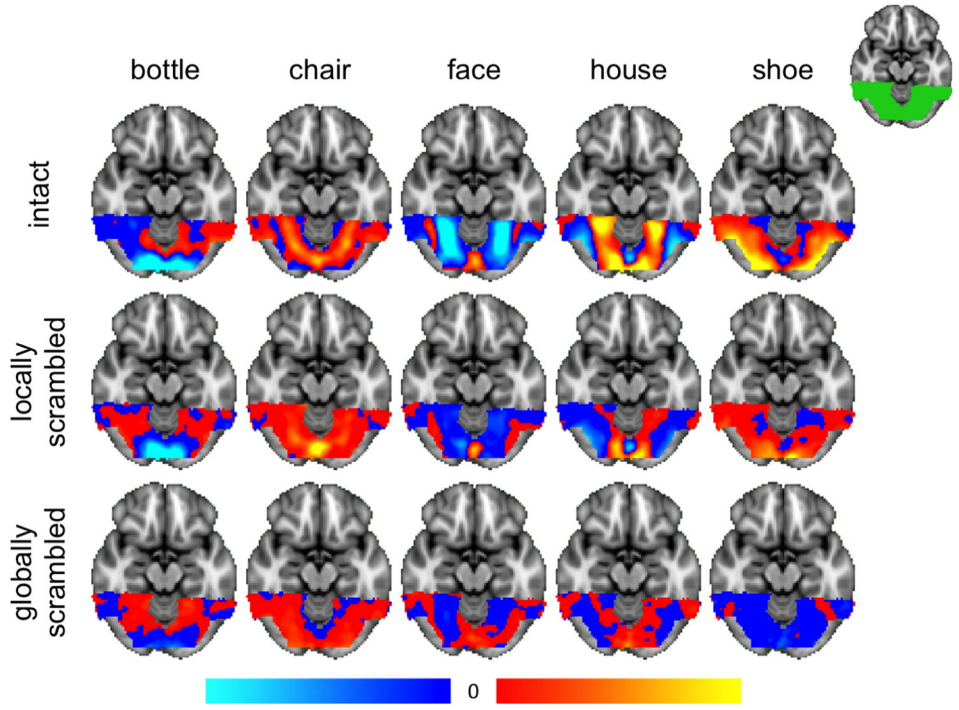

Fig. 3. Patterns of response to different object categories across the ventral visual pathway with intact, locally scrambled and globally scrambled conditions. Red/yellow and blue/light blue colours represent positive and negative fMRI responses, respectively, relative to the mean response across all categories. The scale shows the normalized parameter estimates (beta weights) from 0 to 40. Top-right inset shows the mask of the ventral visual pathway. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

chairs ($t(21) = 8.25$, $p < .0001$), faces ($t(21) = 10.82$, $p < .0001$), houses ($t(21) = 14.85$, $p < .0001$) and shoes ($t(21) = 11.82$, $p < .0001$).

For locally scrambled images, there was a main effect of Comparison ($F(1, 21) = 248.16$, $p < .0001$), and an interaction between Comparison and Category ($F(4, 84) = 10.41$, $p < .0001$). Pairwise comparisons revealed significantly higher within-category compared to between-category correlations for bottles ($t(21) = 8.68$, $p < .0001$), chairs ($t(21) = 6.06$, $p = .0001$), faces ($t(21) = 6.72$, $p < .0001$), houses ($t(21) = 9.63$, $p < .0001$) and shoes ($t(21) = 6.77$, $p < .0001$).

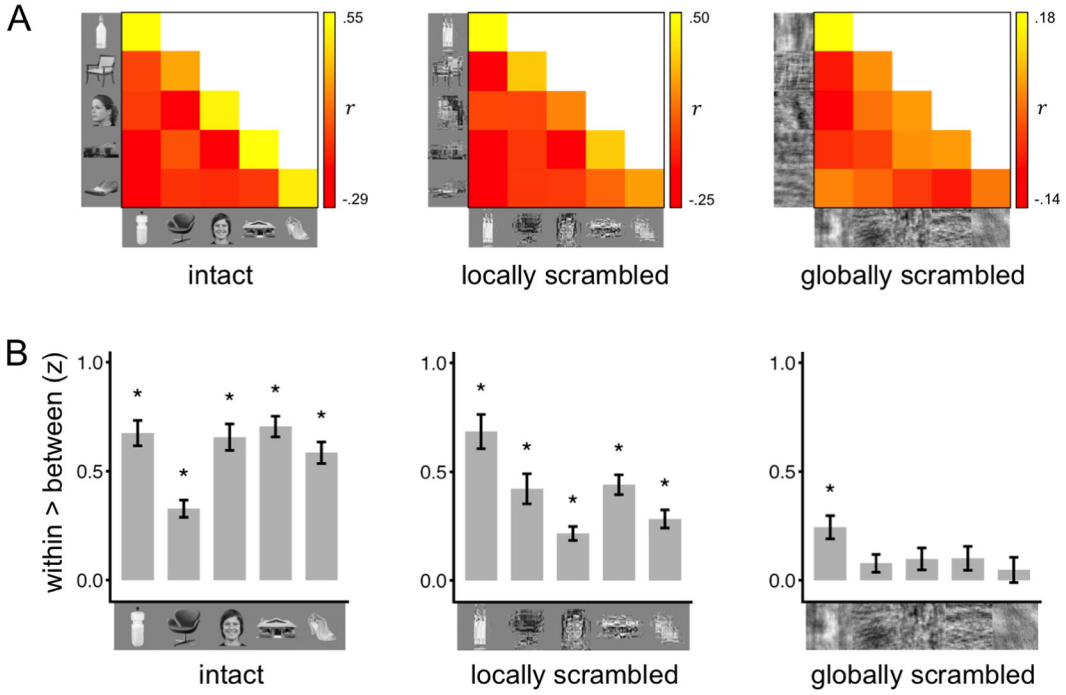

Fig. 4. Distinct patterns of response to different object categories with intact and scrambled images. (A) Similarity matrices showing within- and between-category correlations of neural patterns of response to intact, locally scrambled and globally scrambled object categories. (B) Bar graphs showing within-category minus between-category correlations. Error bars represent ±1 standard error of the mean. * denotes p $_amp_$lt; .05.

For globally scrambled images, there was a main effect of Comparison ($F(1, 21) = 25.00$, $p < .0001$), due to higher within-category compared to between-category correlations. There was no significant interaction between Comparison and Category ($F(4, 84) = 2.09$, $p = .089$). Pairwise comparisons revealed significantly higher within-category compared to between-category correlations for bottles ($t(21) = 4.57$, $p = .0071$), but not for the other object types.

Next, we asked whether the patterns of neural response from intact images were similar to the patterns of response from scrambled images at the group level. We tested this by correlating the group mean similarity matrices for intact versus locally scrambled and intact versus globally scrambled (see Fig. 4A). Scatter plots for each comparison are shown in Fig. 5. The matrix for the intact condition positively correlated with the matrices for both the locally scrambled condition ($r(13) = .88$, $p < .0001$) and the globally scrambled condition ($r(13) = .53$, $p = .043$). This suggests a strong link between responses to intact and scrambled object categories — in particular, the locally scrambled objects.

To determine whether the patterns of neural response from intact images were similar to the patterns of response from scrambled images at the individual level, each participants' locally scrambled or globally-scrambled matrix was correlated with the group mean intact matrix. Both distributions were then contrasted against zero in a one-sample t-test to see if responses to either scrambled image type significantly predicted responses to intact images. Correlations between the group intact matrix and the individual locally-scrambled matrices (mean = .70, SD = .18) were significantly above zero ($t(21) = 18.16$, $p < .0001$). Similarly, correlations between the group intact matrix and the individual globally-scrambled matrices (mean = .23, SD = .27) were significantly above zero ($t(21) = 4.00$, $p = .0007$). Correlations were significantly higher between intact and locally-scrambled matrices than between intact and globally-scrambled matrices ($t(21) = 6.88$, $p < .0001$). This suggests that responses to intact images were better predicted by responses to locally scrambled images than by responses to globally scrambled images.

We then asked whether the explainable variance in intact responses was fully accounted for by the responses to scrambled images, given the level of noise in the data. This was achieved by calculating a noise ceiling (Nili et al., 2014) by taking the mean correlation between each participant's intact similarity matrix and the group mean intact similarity matrix (z = .92). This reflects the maximum similarity that could be expected for any correlation between the intact and locally scrambled conditions. A two-sample, repeated measures t-test revealed that responses to intact images were predicted significantly better by responses to locally scrambled, than globally scrambled images ($t(21) = 6.88$, $p < .0001$). One-sample t-tests revealed that both locally scrambled ($t(21) = 5.71$, $p < .0001$) and globally scrambled ($t(21) = 12.01$, $p < .0001$) correlations with the intact matrix were significantly lower than the noise ceiling (Fig. 6). This analysis suggest that differences between responses to intact and scrambled images could not be entirely accounted for by noise and that other sources of variance are necessary to fully explain the patterns of response to intact images.

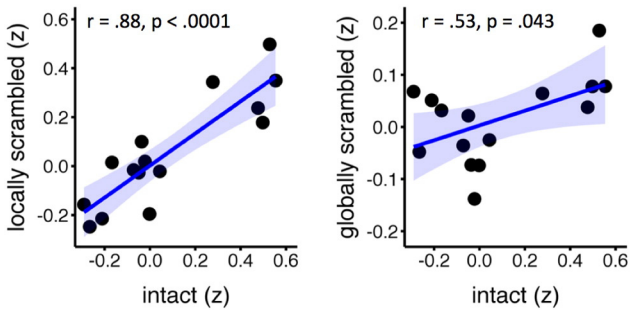

Fig. 5. Similar patterns of response to intact and scrambled images. Scatter plots show the correlation between the similarity matrices in Fig. 4A. Significant positive correlations were found between matrices for intact and each of the scrambled image types, particularly the locally scrambled condition. Shaded regions indicate 95% confidence intervals for the best-fit regression lines.

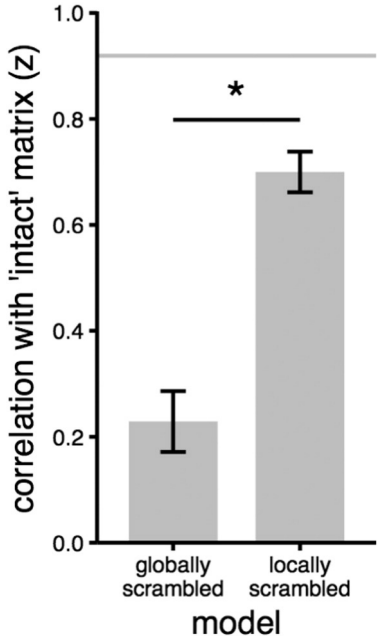

Fig. 6. Bar graph showing how variance in intact responses was accounted for by locally and globally scrambled responses. Grey line represents noise ceiling, which estimates the explainable variance in a dataset given the noise across participants. Responses to intact images were significantly better predicted by responses to locally scrambled images than globally scrambled images. However, neither scrambled condition was able to account for all of the explainable variance. Error bars represent ±1 standard error of the mean. *$p < .0001$.

To determine how the patterns of response might change along the posterior–anterior axis of the ventral pathway, the main anatomical mask was split into 4 subdivisions spanning the posterior to anterior extent of the mask (Supplementary Fig. 2). The MVPA and RSA analyses were then repeated for each slice independently. Again, we found significant correlations between the intact and scrambled matrices. Intact and locally scrambled matrices were significantly correlated in all subdivisions (subdivision 1: $r(13) = .86$, $p < .0005$; subdivision 2: $r(13) = .83$, $p < .0005$; subdivision 3: $r(13) = .92$, $p < .0005$; subdivision 4: $r(13) = .62$, $p < .05$). We also found significant correlations between the intact and globally scrambled matrices in some subdivisions (subdivision 1: $r(13) = .55$, $p < .05$; subdivision 2: $r(13) = .34$, ns; subdivision 3: $r(13) = .54$, $p < .05$; subdivision 4: $r(13) = .44$, ns). We found that category-selective patterns of neural response to all three image types were evident in all subdivisions (Supplementary Table 1), as indicated by a significant effect of Comparison (within-category < between-category correlations).

It is possible, however, that this result could be driven by early visual regions that overlap our ventral stream mask. The retinotopic organization of these regions would likely give rise to similar responses across intact and locally scrambled versions of the same object, as the two images would share a similar spatial envelope. To exclude the possibility that early visual areas were driving our results, we repeated the analysis removing any voxels in our ventral mask that overlapped with early visual areas (V1–V4). To do this, we used the probabilistic maps defined by Wang et al. (2015). Again, we found significant correlations between the intact and scrambled matrices. Intact and locally scrambled matrices were significantly correlated ($r(13) = .85$, $p < .0001$), as were the intact and globally scrambled matrices ($r(13) = .55$, $p < .05$). This suggests that the similarity between responses to intact and locally scrambled images is not dependent on the inclusion of voxels from early visual

regions. We also found category-selective patterns of neural response to all three image types (see Supplementary Table 1).

Recent studies have shown retinotopic organization in more anterior regions of the ventral visual pathway (Wandell et al., 2007; Arcaro et al., 2009; Wang et al., 2015). To investigate the possibility that any regions with retinotopic organization were driving our results, we repeated our analysis removing any voxels in our ventral mask that overlapped with retinotopic regions (V1–V4, V01/2, PHC1/2). Again, we found significant correlations between the intact and scrambled matrices. Intact and locally scrambled matrices were significantly correlated ($r(13) = .86$, $p < .0005$), as were the intact and globally scrambled matrices ($r(13) = .58$, $p < .05$). This shows that the similarity between responses to intact and locally scrambled images is not dependent on the inclusion of voxels with retinotopic visual field maps. We also found category-selective patterns of neural response to all three image types (see Supplementary Table 1).

Although the correlation between matrices from different image types strongly suggests that intact and scrambled images elicit similar patterns of response in the ventral stream, it is possible that strongly related correlation matrices could arise from consistent, but different neural patterns of response to different object categories. To address this issue, we directly compared patterns of response from intact and each scrambled image type in the MVPA. Our rationale was that if the patterns of response in different image types are similar, interchanging them in the analysis should have little effect on the resulting similarity matrix. We found that the mean similarity matrix generated by the cross-correlation of patterns from intact and locally scrambled images was highly correlated with the original intact ($r(13) = 0.93$, $p < .0001$) and locally scrambled ($r(13) = .96$, $p < .0001$) matrices. Additionally, the mean similarity matrix generated by cross-correlation of patterns from intact and globally scrambled images was significantly correlated with the original intact ($r(13) = 0.68$, $p = .0053$) and globally scrambled ($r(13) = .81$, $p < .0005$) matrices. This shows that scrambled images generated similar patterns of response to those evoked by intact images.

We then performed a complementary analysis to estimate the noise ceiling by directly comparing neural patterns of response. The noise ceiling was calculated by measuring the correlation between each participant's response and the remaining group mean response ($z = .48$). The correlation between each participant's responses to locally-scrambled or globally-scrambled images was then compared to the remaining group mean response to intact images. A two-sample, repeated measures t-test revealed that responses to intact images were predicted significantly better by responses to locally scrambled, than globally scrambled images ($t(21) = 6.06$, $p < .0001$). One-sample t-tests revealed that the noise ceiling was significantly higher than the variance explained by both locally scrambled ($t(21) = 13.86$, $p < .0001$) and globally scrambled ($t(21) = 19.92$, $p < .0001$) images. One-sample t-tests also revealed that the variance explained by both locally scrambled ($t(21) = 14.45$, $p < .0001$) and globally scrambled ($t(21) = 4.26$, $p = .0003$) images was significantly above zero.

Next, we asked whether the greater similarity between responses to intact and locally scrambled images, compared to intact and globally scrambled images, could be explained by globally scrambled images evoking less neural activity across the ventral stream. To address this issue, we measured the overall signal change across the ventral stream to each category across the different image conditions (see Fig. 7). A repeated-measures ANOVA was conducted with Image Type and Category as the main factors. There was a main effect of Image Type ($F(2, 42) = 11.03$, $p < .0005$), a marginal effect of Category ($F(4, 84) = 2.51$, $p = .081$) and no interaction between Image Type and Category. Post hoc pairwise comparisons across Image Type revealed that the overall response of the ventral stream was greater for globally scrambled images than for both locally scrambled ($t(21) = 5.25$, $p < .0001$) and intact ($t(21) = 6.00$, $p < .0001$) images. There was no significant difference between the response to locally scrambled and intact images ($t(21) = .043$, ns). This suggests that the greater similarity between intact and locally scrambled images, compared to intact and globally scrambled images, cannot be explained through lower activation of the ventral stream to globally scrambled images.

Finally, we tested the extent to which the temporal position of the red dot within the block affected activity in the ventral visual pathway. It is possible that participants were able to learn that a red dot appeared only once per block, and disengaged with the stimuli after this time. To address this issue, we measured the percent signal change in the ventral stream as a function of the trial number in which the red dot appeared (Supplementary Fig. 1). This was performed separately for each image type. Results were entered into a two-way ANOVA with Image Type and Position (3,4,5,6) as repeated measures. Consistent with the previous analysis, there was a significant main effect of Image Type. However, there was no main effect of Position ($F(3, 63) = 0.52$, ns) and no interaction between Image Type and Position ($F(6, 126) = 2.02$, $p = .067$). This demonstrates that neural activity in the ventral stream did not change based on the temporal position of the red dot trial in a block.

4. Discussion

The aim of the present study was to directly determine whether category-selective patterns of response in the ventral stream were better explained by object category or more basic dimensions of the stimulus. To address this issue, we compared patterns of response to intact and scrambled images. Our hypothesis was that, if category-selective patterns of response reflect the categorical or semantic content of the images, there should be little similarity between the patterns of response elicited by intact and scrambled images. On the other hand, if category-selective patterns are based on more basic image properties, similar patterns should be elicited by both intact and scrambled images. We found distinct and reliable category-selective patterns of response

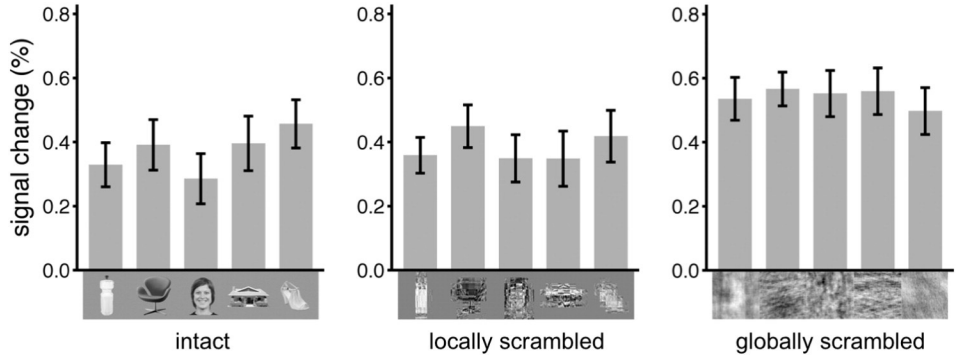

Fig. 7. Average signal change across the ventral stream to each category in each image type. Globally scrambled images evoked more activity than the locally scrambled and intact images. Error bars represent ±1 standard error of the mean.

for both the intact and scrambled image conditions. The patterns of response to intact images were most strongly correlated with the locally scrambled images implying the importance of spatial properties.

Our results show that categorical patterns of response in the ventral visual pathway are evident to images that lack any semantic categorical information. This suggests that the organization of this region is based on more basic properties of the image. These findings are consistent with recent studies in which we have shown that basic image properties of different object categories can predict patterns of response in high-level visual areas (Rice et al., 2014; Watson et al., 2014; Andrews et al., 2015). However, because images drawn from the same category are likely to have similar lower-level properties, it was unclear from previous work whether patterns are determined primarily by membership of a common category or by the shared lower-level image statistics characteristic of that category. The results from the current study directly demonstrate that lower-level properties of the image can predict patterns of response in high-level visual cortex.

Although lower-level image properties account for the majority of the variance in responses to intact images, there remains a significant amount of variance to be explained. According to our noise ceiling estimate, noise only accounts for a portion of this unexplained variance. The remaining variance in the responses to intact images could reflect higher-level or categorical factors as has been proposed in previous studies (Kriegeskorte et al., 2008; Naselaris et al., 2009; Connolly et al., 2012). However, it is also possible that it reflects sensitivity to image properties that are disrupted by either scrambling process. 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 images. The scrambling procedure disrupts many of the statistical relationships between the elements. So, it is possible that image manipulations that can preserve these mid-level properties of objects (cf. Freeman and Simoncelli, 2011) might generate patterns of response that are even more similar to those found for intact objects.

Patterns of response to intact images were more strongly correlated with responses to locally scrambled than globally scrambled images. One key difference between these two conditions is that the spatial properties, such as the shape (or spatial envelope) of the image, are somewhat preserved in the locally scrambled images, but not in the globally scrambled images. The greater similarity between responses to intact locally scrambled images is consistent with previous studies that have shown a modulatory effect of spatial properties on patterns of response in the ventral visual pathway (Levy et al., 2001; Hasson et al., 2002; Golomb and Kanwisher, 2012). Indeed, a number of studies have investigated the sensitivity of the ventral stream to shape information (Drucker and Aguirre, 2009; Haushofer et al., 2008; Kayaert et al., 2003; Kourtzi and Kanwisher, 2001; Op de Beeck et al., 2001; Cant and Xu, 2012; Watson et al., 2016; Bracci and Op de Beeck, 2016). However, it has not been clear whether spatial properties represent a fundamental organizing principle in the organization of the ventral visual stream or whether they just reflect a modification of the underlying category-selective representation due to the statistics of natural viewing (Kanwisher, 2001). The absence of categorical information in the locally scrambled images in this study clearly shows the fundamental importance of spatial properties in the patterns of response.

To determine whether the influence of image properties varied across the ventral stream, we repeated our analysis along different axes of our mask. The patterns of response to locally-scrambled and intact images were significant across all anterior–posterior subdivisions of the ventral stream. We also asked whether the patterns of response to locally-scrambled images could be explained by regions that contain visual field maps (Wandell et al., 2007; Arcaro et al., 2009; Silson et al., 2015; Wang et al., 2015). To address this we removed all regions that showed retinotopic regions from the analysis. Despite removing these regions, we found very similar patterns of response between intact and scrambled images.

The organization of the ventral visual pathway along lower-level dimensions of the stimulus raises the important question about which regions of the brain underpin our ability to make categorical judgments. It is possible that these decisions are based on representations in more anterior regions of the temporal lobe. Indeed, damage to these regions is known to affect categorical perception (Warrington, 1975; Hodges et al., 1992). However, it is not clear whether a representation based on more basic dimensions of the image precludes a causal role in object categorization. Clearly, high-level visual object representations must be constructed from lower-level representations. Indeed, images from the same object category are likely to have similar low-level image properties. So, it is possible that categorical decisions could still involve ventral visual pathway.

Another important feature of our results is that the patterns of fMRI response generalize across participants. fMRI studies have shown that the location of category-selective regions in the ventral visual pathway is broadly consistent across individuals (Kanwisher, 2010). This implies that common principles underpin the organization of this region. 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 (Poldrack et al., 2009; Rice et al., 2014; Watson et al., 2014). Our data show that the patterns of response to different object categories were consistent across individuals and thus reflect the operation of consistent organizing principles.

In conclusion, the findings from these studies provide a new framework in which to consider the organization of high-level visual cortex. Previous attempts to characterize the organization of visual cortex beyond the early stages of visual processing have needed to include categorical or semantic information about the images. However it has never been clear whether this selectivity is driven solely by tunings to discrete object categories or whether it reflects sensitivity to lower-level features that are common to images from a particular category. Here, we show that similar patterns of response are evident to intact objects and scrambled objects that contain similar lower-level properties, but convey negligible categorical information. This suggests that the organization of the ventral visual pathway reflects tuning for more basic properties of the stimulus. With this lower level framework of stimulus representation, it is more straightforward to determine how a continuous map that underpins the perception of objects could emerge (Andrews et al., 2015).

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2016.04.060.

References

Andrews, T.J., Watson, D.M., Rice, G.E., Hartley, T., 2015. Low-level image properties of natural images predict topographic patterns of neural response in the ventral visual pathway. J. Vis. 15 (7), 3.

Arcaro, M.J., McMains, S.A., Singer, B.D., Kastner, S., 2009. Retinotopic organization of human ventral visual cortex. J. Neurosci. 29 (34), 10638–10652.

Bonhoeffer, T., Grinvald, A., 1991. Iso-orientation domains in cat visual cortex are arranged in pinwheel-like patterns. Nature 353, 429–431.

Bracci, S., Op de Beeck, H., 2016. Dissociations and associations between shape and category representations in the two visual pathways. J. Neurosci. 36 (2), 432–444.

Cant, J.S., Xu, Y., 2012. Object ensemble processing in human anterior-medial ventral visual cortex. J. Neurosci. 32 (7885–7700).

Chao, L.L., Haxby, J.V., Martin, A., 1999. Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. Nat. Neurosci. 2, 913–919.

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 participants and posterior split-brain patients. Brain 123, 291–307.

Connolly, A.C., Guntupalli, J.S., Gors, J., Hanke, M., Halchenko, Y.O., Wu, Y.-C., Abdi, H., Haxby, J.V., 2012. The representation of biological classes in the human brain. J. Neurosci. 32, 2608–2618.

Downing, P.E., Jiang, Y., Shuman, M., Kanwisher, N., 2001. A cortical area selective for visual processing of the human body. Science 293, 2470–2473.
Downing, P.E., Chan, A.W.-Y., Peelen, M.V., Dodds, C.M., Kanwisher, N., 2006. Domain specificity in visual cortex. Cereb. Cortex 16, 1453–1461.
Drucker, D.M., Aguirre, G.K., 2009. Different spatial scales of shape similarity representation in lateral and ventral LOC. Cereb. Cortex 19, 2269–2280.
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, 598–601.
Freeman, J., Simoncelli, E.P., 2011. Metamers of the ventral stream. Nat. Neurosci. 14, 1195–1201.
Golomb, J.D., Kanwisher, N., 2012. Retinotopic memory is more precise than spatiotopic memory. Proc. Natl. Acad. Sci. U. S. A. 109, 1796–1801.
Grill-Spector, K., Weiner, K.S., 2014. The functional architecture of the ventral temporal cortex and its role in categorization. Nat. Rev. Neurosci. 15, 536–548.
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, 37–53.
Hasson, U., Levy, I., Behrmann, M., Hendler, T., Malach, R., 2002. Eccentricity bias as an organizing principle for human high-order object areas. Neuron 34, 479–490.
Haushofer, J., Livingstone, M.S., Kanwisher, N., 2008. Multivariate patterns in object-selective cortex dissociate perceptual and physical shape similarity. PLoS Biol. 6, e187.
Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P., 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293, 2425–2430.
Haxby, J.V., Connolly, A.C., Guntupalli, J.S., 2014. Decoding neural representational spaces using multivariate pattern analysis. Annu. Rev. Neurosci. 37, 435–456.
Hodges, J.R., Patterson, K., Oxbury, S., Funnell, E., 1992. Semantic dementia: progressive fluent aphasia with temporal lobe atrophy. Brain 115, 1783–1806.
Hubel, D.H., Wiesel, T.N., 1968. Receptive fields and functional architecture of monkey striate cortex. J. Physiol. 195, 215–243.
Ishai, A., Ungerleider, L.G., Martin, A., Schouten, J.L., Haxby, J.V., 1999. Distributed representation of objects in the human ventral visual pathway. Proc. Natl. Acad. Sci. U. S. A. 96, 9379–9384.
Kanwisher, N., 2001. Faces and places: of central (and peripheral) interest. Nat. Neurosci. 4, 455–456.
Kanwisher, N., 2010. Functional specificity in the human brain: a window into the functional architecture of the mind. Proc. Natl. Acad. Sci. U. S. A. 107, 11163–11170.
Kanwisher, N., McDermott, J., Chun, M.M., 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. J. Neurosci. 17, 4302–4311.
Kayaert, G., Biederman, I., Vogels, R., 2003. Shape tuning in macaque inferior temporal cortex. J. Neurosci. 23, 3016–3027.
Konkle, T., Oliva, A., 2012. A real-world size organization of object responses in occipitotemporal cortex. Neuron 74, 1114–1124.
Kourtzi, Z., Kanwisher, N., 2001. Representation of perceived object shape by the human lateral occipital cortex. Science 293, 1506–1509.
Kriegeskorte, N., Mur, M., Ruff, D.A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., Bandettini, P.A., 2008. Matching categorical object representations in inferior temporal cortex of man and monkey. Neuron 60, 1126–1141.
Levy, I., Hasson, U., Avidan, G., Hendler, T., Malach, R., 2001. Center — periphery organization of human object areas. Nat. Neurosci. 4, 533–539.
McCarthy, G., Puce, A., Gore, J.C., Truett, A., 1997. Face-specific processing in the human fusiform gyrus. J. Cogn. Neurosci. 9, 605–610.
McNeil, J.E., Warrington, E.K., 1993. Prosopagnosia: a face-specific disorder. Q. J. Exp. Psychol. Sect. A 46, 1–10.
Milner, A.D., Goodale, M.A., 1995. The Visual Brain in Action. Oxford University Press, Oxford.
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. J. Cogn. Neurosci. 9, 555–604.
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, 902–915.
Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., 2014. A toolbox for representational similarity analysis. PLoS Comput. Biol. 10, e10003553.
Op de Beeck, H.P., Wagemans, J., Vogels, R., 2001. Inferotemporal neurons represent low-dimensional configurations of parameterized shapes. Nat. Neurosci. 4, 1244–1252.
Op de Beeck, H.P., Haushofer, J., Kanwisher, N.G., 2008. Interpreting fMRI data: maps, modules and dimensions. Nat. Rev. Neurosci. 9, 123–135.
Poldrack, R.A., Halchenko, Y.O., Hanson, S.J., 2009. Decoding the large-scale structure of brain function by classifying mental states across individuals. Psychol. Sci. 20, 1364–1372.
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. J. Neurosci. 34, 8837–8844.
Silson, E.H., Chan, A.W.-Y., Reynolds, R.C., Kravitz, D.J., Baker, C.I., 2015. A retinotopic basis for the division of high-level scene processing between lateral and ventral human occipitotemporal cortex. J. Neurosci. 35, 11921–11935.
Ungerleider, L.G., Mishkin, M., 1982. Two Cortical Visual Systems. In: Ingle, D.J., Goodale, M.A., Mansfield, R.J.W. (Eds.), Analysis of Visual Behavior. MIT Press, Cambridge, MA, pp. 549–586.
Wandell, B.A., Dumoulin, S.O., Brewer, A.A., 2007. Visual field maps in human cortex. Neuron 56, 366–383.
Wang, L., Mruczek, R.E.B., Arcaro, M.J., Kastner, S., 2015. Probabilistic maps of visual topography in human cortex. Cereb. Cortex 25, 3911–3931.
Warrington, E.K., 1975. The selective impairment of semantic memory. Q. J. Exp. Psychol. 27, 635–657.
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, 173–183.