The Neural Representation of Objects in Visual Cortex

David Douglas Coggan

Doctor of Philosophy

University of York

Psychology

January 2019

Abstract

Neuroimaging studies have shown that different categories of object evoke different neural responses in the ventral visual pathway. This has been interpreted to suggest that these regions represent high-level conceptual or semantic properties of the stimulus, such as its category. However, images from different categories differ in low-level visual properties. Therefore, the extent to which category-specific neural responses indicate high-level or low-level representations is unclear. This thesis investigates the extent to which low-level properties of objects are important in the neural response of ventral visual pathway. The first study uses a data-driven approach to select clusters of objects based on the similarity of their low-level visual properties. These visually defined clusters did not correspond to typical object categories, but still evoked distinct patterns of response in the ventral stream. The second and third studies show category-specific patterns of response in the ventral stream to scrambled objects that are not recognizable, but nevertheless retain many of their low-level visual properties. The fourth study reveals that the bias toward natural object images found in the ventral stream begins to emerge in early visual areas. The final chapter shows that category-specific patterns of EEG response can be also explained by low-level image properties. Taken together, these results demonstrate the importance of low-level visual properties in the neural representation of objects. These findings suggest that the category-selectivity observed in high-level visual regions can be explained by a distributed organization based around more basic properties of the stimulus.

List of Contents

List of Tables

List of Figures

Acknowledgements

First of all, I would like to thank my supervisors, Professor Tim Andrews and Dr Daniel Baker, from whom I have learned so much. I know the two of you had many responsibilities besides supporting me, but it never felt that way. I cannot overstate my gratitude to you both.

I would also like to thank the other members of my panel, Dr Tom Hartley and Dr Shirley-Ann Rueschemeyer. Your enthusiasm and insight made for very enjoyable meetings over the last few years.

Special mention also goes to Dr David Watson and Dr André Gouws. Thank you for sharing your technical expertise and, more importantly, your code.

To my house-mate, Holly Brown, thank you for tolerating me over the years; for the cakes you so frequently plied me with and for making our house feel like home.

Finally, none of this would have been possible without my parents. Your moral and financial support have been critical at times and I cannot express how lucky I feel to have such a wonderful family behind me.

To the memory of Dorothy Coggan (1920 – 2018)

Author’s Declaration

This thesis presents original work completed by the author, David Coggan, under the joint supervision of Prof. Timothy Andrews and Dr. Daniel Baker. This work has not previously been presented for an award at this, or any other, university. All sources are acknowledged as references.

The empirical work presented in this thesis has been published or is currently under review in the following peer-reviewed journals:

Coggan, D. D., Watson, D. M., Hartley, T., Baker, D. H., & Andrews, T. J. (under review). A data-driven approach to stimulus selection reveals the emergence of an image-based representation of objects in high-level visual areas. *Cerebral Cortex.*

Coggan, D. D., Baker, D. H., & Andrews, T. J. (in press). Selectivity for mid-level properties of faces and places in the fusiform face area and parahippocampal place area. *European Journal of Neuroscience.*

Coggan, D. D., Allen, L. A., Farrar, O. R. H., Gouws, A. D., Morland, A. B., Baker, D. H., & Andrews, T. J. (2017). The emergence of object-selectivity in early visual areas (V1-V3). *Scientific Reports, 7,* 2444.

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*, 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.

Results from empirical chapters have been presented at the following conferences:

Talks

Coggan, D. D., Watson, D. M., Hartley, T., Baker, D. H., & Andrews, T. J. (2017). A data-driven approach to stimulus selection reveals the importance of visual properties in the neural representation of objects. *Vision Sciences Society, Florida, USA.*

Coggan, D. D., Watson, D. M., Baker, D.H., Hartley, T., & Andrews, T. J. (2017). The importance of visual properties in the emergence of higher-level representations in the ventral visual pathway. *Experimental Psychological Society, London, UK.*

Coggan, D. D., Baker, D. H., & Andrews, T. J. (2016). Investigating the temporal properties of visual object processing using a multivariate analysis of EEG data. *Vision Sciences Society, Florida, USA.*

Posters

Coggan, D. D., Watson, D. M., Hartley, T., Baker, D. H., & Andrews, T. J. (2018). A data-driven approach to stimulus selection reveals an image-based representation of objects in high-level visual areas. *Applied Vision Association, Bradford, UK.*

Coggan, D. D., Liu, W., Baker, D. H., & Andrews, T. J. (2015). Category-selective patterns of neural response to scrambled images in the ventral visual pathway *Vision Sciences Society, Florida, USA.*

Chapter 1. Literature Review

The ability to perceive and extract information from objects in the environment is critical to the survival of many species, for instance when locating food, avoiding predators, or searching for conspecifics. Human object perception appears to be particularly geared toward categorisation - subjects asked to divide an array of objects into groups based on perceptual similarity will tend to group them by category (Edelman, 1998). Object categorization occurs automatically (Grill-Spector & Kanwisher, 2005) and rapidly (within 200 ms; Grill-Spector & Kanwisher, 2005; Potter, 1976; Thorpe, Fize, & Marlot, 1996), even under highly degraded viewing conditions (Ullman, 1998; Wyatte, Curran, & O'Reilly, 2012). This performance is especially impressive given that object categorisation is a highly computationally demanding process, largely due to the variability in viewing conditions. For example, a single object can generate a virtually infinite number of different retinal projections based on many factors, for instance its location, illumination, or viewpoint. Moreover, objects from the same category may vary in colour, size, texture and other features. The reliable identification or categorisation of a target object therefore requires numerous transformations of the raw visual input in order to match it to a stored object representation, for instance a structural description of its parts and their relations (Biederman, 1987). Solving object perception consequently requires a large amount of computational resource and a complex cognitive architecture. Indeed, artificial visual systems could not achieve human-level performance on object categorisation tasks until recently (Kriegeskorte, 2015; Szegedy et al., 2015).

The brain regions involved in object processing are found in the occipital and temporal lobe. Sensory signals from the retina travel via the lateral geniculate nucleus to the primary visual cortex (V1) at the occipital pole. Here, neighbouring neurons receive input from neighbouring regions of the retina, such that the region as a whole contains a topographic map of the visual field (Hubel & Wiesel, 1968). From here, signals are relayed through a series of contiguous visual regions along the cortical surface in two processing streams - one projecting superiorly toward parietal cortex and the other projecting anteriorly into ventral temporal cortex (Milner & Goodale, 1995, 2008; Ungerleider & Mishkin, 1982). The latter stream, known as the ventral visual pathway, has been shown by neuropsychological studies to be critical for object perception. A lesion in this brain region often results in impaired perception and recognition of different categories of object, depending on the location and extent of the damage. For example, damage to the fusiform gyrus or ventral occipital cortex can disrupt face recognition – a condition known as prosopagnosia (Bouvier & Engel, 2006; De Renzi, Perani, Carlesimo, Silveri, & Fazio, 1994; McNeil & Warrington, 1993). Damage to other regions within the ventral stream can impact object (Farah, 1990; Moscovitch et al., 1997) or scene (Mendez & Cherrier, 2003) perception while relatively preserving face perception. These case studies demonstrate that the ventral stream is causally involved in our capacity for object perception.

Consistent with these neuropsychological reports, functional magnetic resonance imaging (fMRI) studies have shown that discrete regions of the ventral visual pathway are apparently specialized for different categories of objects. For example, a region within the fusiform gyrus referred to as the fusiform face area (FFA) shows consistently higher neural activity in response to images of faces than to images of non-face objects (Kanwisher et al., 1997; McCarthy et al., 1997). Selectivity for faces has also been revealed in a region of inferior occipital gyrus known as the occipital face area (OFA; Clark et al., 1996; Gauthier et al., 2000). Other category-selective modules have been located in neighbouring cortical regions, including the parahippocampal place area (PPA), which responds highest to images of scenes or buildings (Epstein & Kanwisher, 1998); the extrastriate body area (EBA) which is selective for images of body parts (Downing, Jiang, Shuman, & Kanwisher, 2001); and the visual word-form area, which is activated by visually-presented words (Cohen et al., 2000). These regions also exhibit category-selectivity through fMR-adaptation (Grill-Spector & Malach, 2001), in which the response to a visually presented object gradually diminishes if the presentation is prolonged or repeated. The FFA, PPA and EBA have shown adaptation to their preferred categories, but not to their non-preferred categories (Andrews, Clarke, Pell, & Hartley, 2010; Myers & Sowden, 2008). These findings are further complemented by studies in which the function of category-selective regions is disrupted using transcranial magnetic stimulation (TMS) or direct cortical stimulation. Targeting these regions selectively impacts the processing of their preferred category, suggesting that they play a causal role (Parvizi et al., 2012; Pitcher, Charles, Devlin, Walsh, & Duchaine, 2009; Rangarajan et al., 2014; Sadeh et al., 2011). In conjunction with the neuropsychological literature, these studies indicate that different object categories recruit distinct cortical regions within the ventral stream. This has been interpreted in favour of a modular organizing principle in which different regions of cortex selectively process different object categories (Kanwisher, 2010).

Despite the category-selectivity observed in the ventral stream, specialized regions have only been reported for a limited number of categories (Downing, Chan, Peelen, Dodds, & Kanwisher, 2006; Op de Beeck, Haushofer, & Kanwisher, 2008). It is therefore unclear how a modular account could explain the human capacity to recognise a large number of object categories. However, an alternative organising principle has emerged from studies that employ multivariate analyses of the ventral response. Neuroimaging studies discussed thus far have employed univariate techniques, in which category-selectively is determined by analysing each location in the brain independently. A multivariate analysis examines the spatial pattern of response across many voxels simultaneously. Using these techniques, many studies have found that the pattern of response across the ventral stream can distinguish a far greater range of object categories (e.g. Carlson, Schrater, & He, 2003; Cox & Savoy, 2003; Haxby et al., 2001; Hung, Kreiman, Poggio, & DiCarlo, 2005; Ishai, Ungerleider, Martin, Schouten, & Haxby, 1999; Kiani, Esteky, Mirpour, & Tanaka, 2007; Kriegeskorte, Mur, Ruff, et al., 2008; Spiridon & Kanwisher, 2002; Tsao, Freiwald, Tootell, & Livingstone, 2006). The pattern of response remains diagnostic of the object category even when discrete, category-selective regions are excluded from the analysis (Haxby et al., 2001). In contrast to a modular organization, the distributed nature of the response has been interpreted as showing a topographic map of object form (Haxby et al., 2001).

The topographic organization proposed by Haxby and colleagues (2001) is thought to be analogous to those found in early visual areas, in which low-level properties of an image, such as spatial frequency, orientation and retinal position, are mapped continuously across the cortical sheet (Bonhoeffer & Grinvald, 1991; Engel et al., 1994; Hubel & Wiesel, 1968; Wandell, Dumoulin, & Brewer, 2007). In these studies, the neural response is monitored while a stimulus gradually traverses along its basic functional dimension. If the locus of neural activation shifts smoothly across the cortical surface, a map has been identified. For example, a stimulus that drifts across the visual field evokes a smoothly transitioning wave of neural activity across the primary visual cortex (Wandell et al., 2007). Conversely, if the activation jumps from one location to another, this would constitute evidence for discrete modules.

Although mapping techniques have proved useful in identifying maps in auditory (Humphries, Liebenthal, & Binder, 2010) and early visual (Wandell et al., 2007) cortices, two main issues arise when it is applied to object categories in the ventral stream. First, object category is a complex, multi-dimensional feature space that is difficult to parameterize. For instance, an object can be categorised at different levels of abstraction, ranging from subordinate (e.g. Labrador) to basic (e.g. dog) to super-ordinate (e.g.

animal). Objects may also have different functional and sensory properties that obfuscate its category, for instance a toy car is used as a toy but has the appearance of a car. One solution is to use human perceptual judgments to parameterise object category, since human perception is tightly linked to categorical properties (Edelman, 1998; Nosofsky, 1986). For example, subjects can rate the perceptual similarity between each pairing of objects in a stimulus set, thereby producing model of categorical relations against which neural responses can be compared (e.g. Edelman, Grill-Spector, Kushnir, & Malach, 1998; Haushofer, Livingstone, & Kanwisher, 2008). However, this represents only a partial solution as an independent parameterisation of object category is not obtained. A second, more general problem with a map-like representation in the ventral stream is that object category is a discontinuous variable. Continuous changes are often possible within a category as corresponding features can be mapped across exemplars, for instance when morphing between faces. However, problems arise when crossing categorical boundaries, for instance from face to body part. In summary, the complex and discontinuous nature of object category has prevented its description along a single continuous parameter. Consequently, whether category-selective regions constitute independent specialized modules or districts of a larger continuous map is still a matter of current debate (Haxby, Connolly, & Guntupalli, 2014; Kanwisher, 2010).

While category-selective responses in the ventral stream have been interpreted to suggest that the fundamental stimulus dimension along which the neural responses are organised is object category (Connolly et al., 2012; Naselaris, Prenger, Kay, Oliver, & Gallant, 2009), other high-level dimensions have been proposed. For example, one study measured patterns of ventral response to object images that spanned a range of different categories including faces, body parts, and everyday objects that were natural or artificial (Kriegeskorte, Mur, Ruff, et al., 2008). Responses to each object were projected into a two-dimensional feature space in which nearby objects evoked similar responses. Objects that were animate (faces, body parts) and inanimate (objects) occupied different halves of this feature space, suggesting that animacy might represent a key representational dimension. Another study hypothesised that responses might be organised around an object's real-world size, as this feature heavily constrains how humans can act upon it (Konkle & Oliva, 2012). Objects of small and large real-world size were presented at the same retinal size yet could be differentiated through patterns of response in the ventral stream, thereby supporting their hypothesis. Other investigations have shown ventral responses to semantic or conceptual knowledge associated with object images (Martin et al., 1996; Chao et al., 1999; Beauchamp et al., 2002; Mahon et al., 2009), indicating that the ventral stream may support amodal object representations. However, it is difficult to determine from these studies which dimension might represent the fundamental organizing principle of the ventral stream, as they are interdependent. For example, the animate / inanimate dimension could be considered a categorical distinction at a superordinate level (Grill-Spector & Weiner, 2014). Additionally, objects from the same category are likely to be of similar real-world size and share conceptual attributes. Consequently, it is not clear which, if any, of these high-level dimensions represents the basic response tuning of the ventral stream. Nevertheless, prevailing opinion is that it is some high-level property or properties (Grill-Spector & Weiner, 2014; Haxby et al., 2014; Kanwisher, 2010; Kriegeskorte, Mur, Ruff, et al., 2008; Op de Beeck et al., 2008).

An alternative possibility is that category-selective responses are driven by a tuning to more basic visual properties that covary with categorical properties (Andrews, Watson, Rice, & Hartley, 2015). Images from the same object category are likely to share visual properties (Honey, Kirchner, & VanRullen, 2008; Lescroart, Stansbury, & Gallant, 2015; O'Toole, Jiang, Abdi, & Haxby, 2005; Rice, Watson, Hartley, & Andrews, 2014). Therefore, category-specific responses could be expected under both high-level and low-level accounts of neural organization. For example, the face-selectivity of a cortical region or neural activation pattern could reflect a tuning to the low-level cues typically found in face images, such as curvature and particular patterns of horizontal orientations (Dakin & Watt, 2009; Valérie Goffaux & Dakin, 2010), rather than the categorical or semantic information that such images convey. A number of studies have shown response biases to basic visual features in the ventral stream. For example, retinotopic responses in category-selective regions have been identified, ranging from eccentricity biases (Uri Hasson, Levy, Behrmann, Hendler, & Malach, 2002; Levy, Hasson, Avidan, Hendler, & Malach, 2001; Weiner et al., 2014) to complete maps of the visual field (Arcaro, McMains, Singer, & Kastner, 2009; Brewer, Liu, Wade, & Wandell, 2005). Furthermore, the PPA has been shown to respond selectively to rectilinear features (Nasr, Echavarria, & Tootell, 2014) and higher spatial frequencies (Rajimehr, Devaney, Bilenko, Young, & Tootell, 2011), which are common in scene images. In addition to strictly low-level properties, mid-level visual features such as shape cues also appear to be important in the ventral stream. For instance, artificial object categories with different shape characteristics evoke different neural responses in face-selective regions (Op de Beeck, Torfs, & Wagemans, 2008). However, while these studies show a sensitivity to visual properties in the ventral stream, the extent to which this might explain high-level effects is not directly tested.

Other studies have examined the role of low-level features in the context of categorical responses by quantifying visual differences between categories in order to model the neural response. For instance, a classifier applied to both computational descriptions of image properties and patterns of neural response showed that pairs of categories that were more visually confusable produced more similar responses (O'Toole et al., 2005). Similarly, patterns of response to different object and scene categories have been linked to the spatial properties of the image (Andrews et al., 2015; Rice et al., 2014; Watson, Hartley, & Andrews, 2014; Watson, Hymers, Hartley, & Andrews, 2016). In these studies, categories with more similar image properties evoked more similar patterns of neural response. In other work, Clarke et al. (2014) modelled different features of an object stimulus set, including category, semantic information and a computational model designed to capture early visual processing. Mapping the unique contributions of each model to the neural response, they found that visual properties predicted responses beyond early visual cortex in parts of the ventral stream. Taken together, these findings suggest that visual properties offer a plausible explanation for the category-selectivity observed in the ventral stream. The general approach of mapping visual differences between categories onto neural data is useful in that an unlimited number of models can be applied to the data without the need to add experimental conditions. However, they suffer from two key limitations. First, they do not experimentally dissociate high- and low-level features. Visual properties are not only more similar across images within a category than between categories, but the visual similarity between different categories is predicted by their semantic similarity (Deselaers & Ferrari, 2011). Therefore, the success of a visual model in predicting the similarity in neural response across categories could be expected under both categorical and visual organizing principles.

Attempts have been made to experimentally distinguish the effects visual and categorical features on ventral responses. For example, a number of studies have presented the same category across a variety low-level transformations, for instance retinal size (Eger, Schyns, & Kleinschmidt, 2004; Grill-Spector et al., 1999; Liu, Agam, Madsen, & Kreimen, 2009; Vuilleumier, Henson, Driver, & Dolan, 2002) and position (Andrews & Ewbank, 2004; Grill-Spector et al., 1999; Liu et al., 2009; MacEvoy & Epstein, 2007; Schwarzlose, Swisher, Dang, & Kanwisher, 2008). In these studies, categorical preferences were consistent across the different transformations and were subsequently interpreted in favour of high-level organising principles. However, other studies using multivariate approaches have found that patterns of response to different categories are indeed reliably modulated by basic transformations, such as retinal size (Watson, Young, & Andrews, 2016) and spatial frequency content (Watson, Hymers, et al., 2016). A limitation of using this approach to establish categorical organising principles is that it does not account for other visual features that are unaffected by these manipulations. For example, an object's shape and the spatial configuration of visual features are unaffected by size and position changes. Such properties likely are important in later stages of the ventral stream, where receptive field sizes become large enough to represent size- and position-invariant visual features (Dumoulin & Wandell, 2008). Consequently, it is unclear whether category effects that are robust to size and position changes are driven by categorical information or visual properties that are preserved across image transformations.

A complementary approach is to remove the high-level information conveyed by object images while preserving visual features. In this process, a control stimulus set is generated that retains the visual properties of each image without expressing high-level cues. Examples of these techniques are illustrated in Figure 1.1 and include box-scrambling, phase-scrambling, texture-scrambling (Portilla & Simoncelli, 2000) and diffeomorphing (Stojanoski & Cusack, 2014), although more coarse forms of spatial rearrangement have also been used (Kanwisher et al., 1997). A number of studies have shown that scrambled images do not produce the category-selective neural responses that emerged using the original images (Allison et al., 1994; Downing et al., 2001; Epstein & Kanwisher, 1998; Kanwisher et al., 1997). This would appear to suggest that the category-selectivity cannot be reduced to visual differences between categories. However, other studies have shown that the selectivity for faces and scenes can be partially replicated using phase-scrambled images (Andrews et al., 2010; Rossion, Hanseeuw, & Dricot, 2012). A persistent issue with this approach is that no single scrambling technique preserves all visual properties. For example, phase-, box- and texture-scrambling destroy visual cues such as shape and the spatial configuration of low-level features. Diffeomorphing better preserves this information, but heavily impacts orientation cues. As any of these properties could be diagnostic of object category, it is possible that category effects that fail to emerge when using scrambled versions of the stimuli could still be driven by image-based differences between categories. However, a useful aspect of this approach is that the transformation often renders the image unrecognisable, thus removing high-level cues. Category effects that are successfully replicated with these images therefore serve as strong evidence that basic visual properties are driving the observed category selectivity (e.g. Andrews et al., 2010; Rossion et al., 2012).

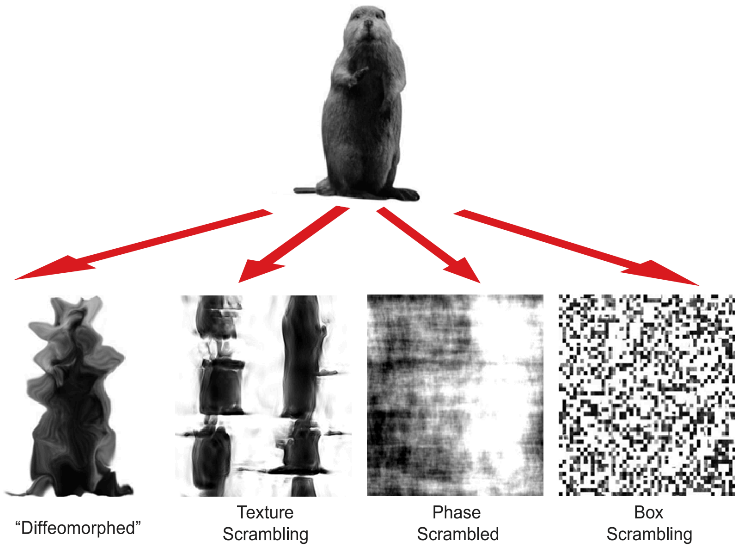

Figure 1.1 *Illustration of different scrambling techniques taken from Stojanoski & Cusack (2014). Each transformation is designed to preserve visual properties while preventing the perception of high-level cues, such as category.*

In summary, distinguishing the category of a visual object appears central to human visual object perception. Supporting brain regions produce neural responses that are tightly linked to categorical properties. However, it is unclear the extent to which these responses can be explained by an underlying tuning to more basic visual features. The development of image-scrambling techniques that better preserve higher-order visual properties may help in this regard. The experiments in this thesis examine the neural response to visual objects in order to establish whether basic visual properties can account for category-selective responses in the ventral stream.

Chapter 2. Methods

2.1. The fMRI BOLD signal

Functional Magnetic Resonance Imaging (fMRI) is a non-invasive technique for measuring neural activity in the brain *in vivo*. When neurons fire, they consume glucose and oxygen, which are delivered via the bloodstream. As the neural firing rate increases, the brain accommodates the greater demand for resources through a local increase in blood-flow. This alters the proportion of oxyhaemoglobin and deoxyhaemoglobin; whose differing magnetic properties drive subsequent changes in the local magnetic field. These changes can be detected through MRI using a contrast referred to as the Blood Oxygenation Level-Dependent (BOLD) response (Ogawa, Lee, Kay, & Tank, 1990). This response lags approximately 6 seconds behind neural firing, thus neural events are difficult to differentiate in time. However, it can be sampled at a high spatial resolution of a few millimetres, depending on factors such as magnetic field strength and desired sample rate. As such, we can infer levels of neural activity across the brain by using fMRI to measure regional vascular changes via the BOLD signal.

2.2. Electroencephalography (EEG)

EEG is another non-invasive method by which neural activity in the brain can be measured *in vivo*. Neurons communicate with one another by releasing neurotransmitters, which generate post-synaptic potentials when they bind to receptor sites on the post-synaptic cell. When a large number of neighbouring neurons fire synchronously, these post-synaptic potentials can sum such that they are detectable using an electrode placed on the scalp. Due to the signal diffusion through the tissues found in the head, these potentials are stronger in sources closer to the electrodes. This can be exploited to coarsely estimate the location of the source by comparing responses across an array of electrodes distributed strategically across the scalp. The key advantage of this technique, however, is that electrodes can sample potentials at each millisecond, providing temporal resolution over three orders of magnitude higher than fMRI. In this way, the time-course of large-scale neural activity in the brain can be directly measured with EEG.

2.3. Univariate Analysis

2.3.1. General Linear Model

A common approach to analysing fMRI data is to use a univariate general linear model (GLM) of the BOLD signal (Friston et al., 1994). Here, a regressor for each stimulus condition is defined in order to model the neural response. For example, Figure 2.1 shows how a box-car model can be used to predict the timeseries of the neural response. First, the model is defined by predicting zero response when a condition is absent and a non-zero response when the condition is present. To account for the temporal properties of the haemodynamic response to neural activity, the model must then be convolved with a canonical haemodynamic response function (HRF). Finally, this model is regressed against the timeseries of the fMRI BOLD signal. This is performed independently for each voxel, with each assigned a regression coefficient or parameter estimate for each condition that reflect the fit of the model. The parameter estimates in each voxel can then be contrasted to establish selectivity for one condition over another condition. These are often represented in a whole-brain statistical map. The statistical significance of these contrast parameter estimates can then be determined either at this stage or after a higher-level analysis across scan sessions and/or subjects. Resulting statistical maps can then be expressed in p-values or z-scores or submitted to further statistical analysis as necessary.

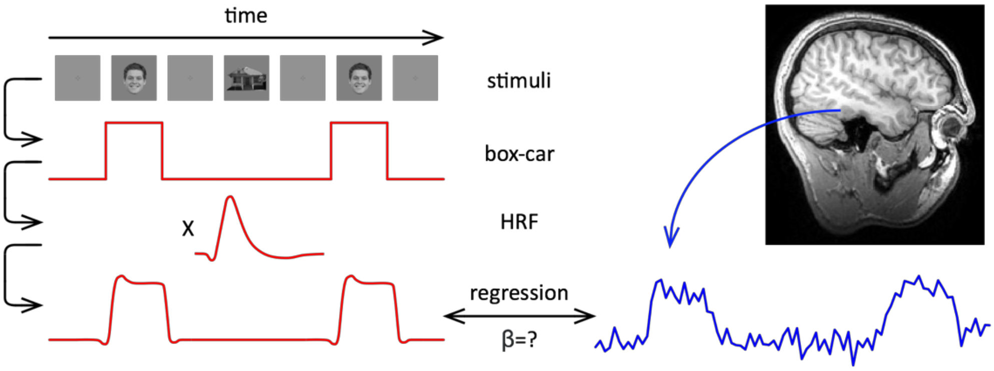

Figure 2.1 *Example of a univariate GLM of fMRI data. First, a box-car function is defined according to periods at which a particular stimulus condition is presented. This function is then convolved with a canonical HRF to produce haemodynamic model, which is then regressed against the timeseries of the BOLD signal for each voxel in the brain. This produces a statistical map of parameter estimates that reflects the responsiveness of each voxel to the stimulus condition.*

2.3.2. FMR-Adaptation

The spatial resolution of fMRI is limited by the size of the voxel. Although reducing voxel size improves spatial resolution, it also reduces the signal in that voxel making it harder to distinguish differences between conditions. Reducing voxel size also reduces the amount of brain that can be measured in a typical fMRI (EPI) pulse sequence. As a compromise, the majority of fMRI experiments have voxels sizes of 2-3 mm^3. However, a voxel of this size still contains several hundred thousand neurons and a standard univariate contrast cannot distinguish the responses of different populations of neurons within the same voxel. For example, it is possible for two stimulus conditions to evoke the same parameter estimate in a voxel through activating entirely separate neural populations of similar size.

FMR-adaptation (Grill-Spector & Malach, 2001; Krekelberg, Boynton, & van Wezel, 2006) is a univariate fMRI experimental paradigm that can measure the sensitivity of individual neural populations to a particular stimulus feature at a sub-voxel level. This approach is based on stimulus repetition effects, whereby repeated or prolonged presentation of the same stimulus produces a reduction in the neural response over time. In humans, this has taken the form of a reduction in BOLD signal to repeated stimuli and has been proposed to stem from neural adaptation – a reduction in the firing rate of a specific neural population activated by the stimulus (Grill-Spector & Malach, 2001). Neural adaptation has been measured directly through single-unit recordings in macaque, in which neurons in inferotemporal cortex (IT) show reduced spiking when face and shape stimuli are repeatedly presented (Li, Miller, & Desimone, 1993; Miller, Li, & Desimone, 1991; Rolls, Baylis, Hasselmo, & Nalwa, 1989; Sobotka & Ringo, 1993). A typical fMR-adaptation experiment will involve contrasting the BOLD response between a sequence of identical stimuli and a sequence in which one stimulus feature is varied. In theory, a voxel containing neurons tuned that stimulus feature will have the same sub-population of neurons repeatedly activated during the former condition, but more different populations activated throughout the latter condition. Adaptation therefore occurs to a greater extent in the first condition and this manifests as a smaller BOLD response in the voxel. In this way, fMR-adaptation can be used to make inferences about the sensitivity of a region of the brain to a specific stimulus feature.

There are, however, a number of limitations with fMR-adaptation that must be taken into consideration (reviewed in Larsson, Solomon, & Kohn, 2015). For, it has been suggested that the level of adaptation may reflect the aggregate of a range of underlying neural effects, including refined tuning, fatigue, response facilitation and altered response dynamics. The relative influence of these effects is unclear and may depend on the experimental paradigm. Furthermore, the finding of fMR-adaptation does not by itself indicate that the region in question is sensitive to the stimulus feature – adaptation in upstream regions has been shown to propagate to later regions. In summary, fMR-adaptation is a useful technique that can measure sensitivity to a stimulus feature, provided that they are designed and interpreted with caution.

2.4. Multivariate/Multivoxel Pattern Analysis (MVPA)

While univariate approaches are powerful tools for fMRI analysis, there are ways in which information is represented in the neural response that cannot be elucidated through these techniques. Both are based on comparing the amplitude of the BOLD response between stimulus conditions on a voxel-by-voxel basis, with statistically significant voxels indicating stimulus-related neural activity. However, it does not follow that voxels yielding sub-threshold responses do not encode information about the stimulus. For example, a stimulus condition can evoke a reliable a pattern of response, distributed across larger regions of the brain. These often include both positive and negative levels of activation and may contain very few voxels showing above threshold activation. Despite this, these patterns can be reliably produced by the stimulus and can vary systematically across conditions, suggesting that information about the stimulus is still encoded in these sub-threshold voxels. By definition, univariate analyses are insensitive to information represented in distributed patterns of response. More recently, multivariate methods have been developed in order to tap into this type of representation.

An alternative to traditional univariate analyses is to examine the spatial patterns of neural response across a number of voxels. Multi-Voxel or Multivariate Pattern Analysis is an umbrella term for these techniques (Haynes, 2015; Schwarzkopf & Rees, 2011). MVPA provides a different sensitivity to the information encoded in neural responses than univariate analyses. This sensitivity is potentially greater as it is often the case that conditions that fail to produce differences in the magnitude of the response can be distinguished on the basis of response patterns. For example, while only three categories of object (faces, scenes, bodies) have been shown to reliably activate a particular region of the ventral stream, many more categories can be distinguished on the basis of the pattern of ventral response (Haxby et al., 2001). The initial goal of MVPA is often to determine whether different stimulus conditions evoked distinct and reliable patterns of response. This can be achieved through either correlation- or classifier-based MVPA.

2.4.1. Correlation-based MVPA

The simplest form of MVPA involves correlation analyses across patterns of response to different conditions, and was the approach used in the earliest multivariate analyses of fMRI data (Haxby et al., 2001; Ishai et al., 1999). In fMRI, patterns of response to each condition are typically constituted by the parameter estimates taken from the GLM. They are then restricted to a subset of voxels constituting a particular region of interest (ROI). To establish whether the patterns of response to each condition are distinct and reliable, the patterns are then cross-validated. Two independent measures of the response to each condition are therefore required. To obtain these, a GLM can be configured to generate separate parameter estimates for odd and even presentations of each condition. This is common practice when running the cross-validation within each subject. Alternatively, responses can be compared across subjects by comparing parameter estimates from one subject to those obtained from a higher-level analysis of the remaining group. This procedure is termed Leave One Participant Out (LOPO), and is the method used in all fMRI multivariate analyses in this thesis. It is possible to enter the restricted parameter estimates into the MVPA directly, however, at this stage the patterns are likely to contain variation that is common to all conditions and therefore unrelated to the key dimensions of the stimulus conditions. Sources of this variation include generic responses to visual stimulation; attentional effects or responsivity differences across voxels arising from biological or neuroimaging factors. To remove this shared variance, patterns can undergo a normalisation procedure. This involves a voxel-wise subtraction of the mean pattern of response across all conditions from the pattern of response to each condition. If desired, these patterns may be further normalised by dividing each voxel's response across the conditions by its standard deviation, thereby generating patterns of z-scores. Normalisation must be applied independently across the data split in order for the two measures to remain independent (Kriegeskorte, Simmons, Bellgowan, & Baker, 2009).

Once patterns have been normalised, they can then be entered into the MVPA. Pattern correlations are obtained for each pairwise permutation of the conditions across the data split, including correlations between responses to the same condition. This process is illustrated in Figure 2.2. As even a small number of conditions can result in many permutations, results are often represented in a correlation matrix. Once this has been performed for all subjects, it is possible to test whether the different conditions can be distinguished on the basis of the neural response. This involves comparing the similarity in responses to the same condition (within-condition) with the similarity in responses across different conditions (between-condition). If within-condition correlations are reliably higher than between-condition correlations in a particular ROI, we can infer that this region is involved in representing the dimension(s) along which our stimulus conditions vary.

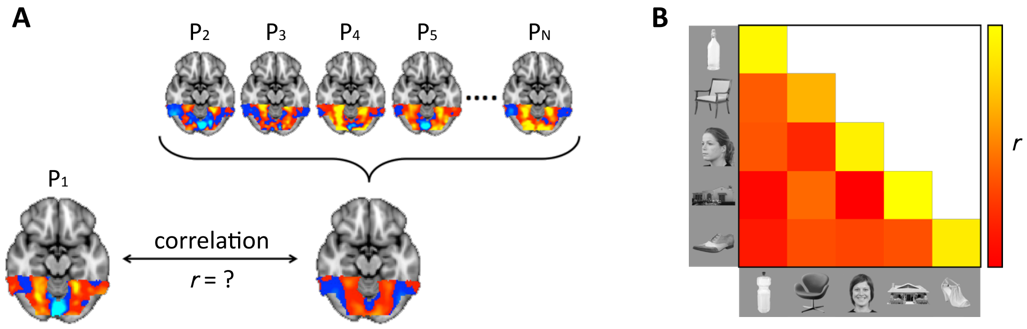

Figure 2.2 *Illustration of correlation-based MVPA. (A) Patterns of response to two stimulus conditions across independent measurements (e.g. individual versus mean of remaining group) are obtained and restricted to a ROI. A Pearson's correlation analysis is then performed on the patterns. This is then repeated for each pairwise permutation of within- and between-condition comparisons. (B) Results can be represented in a correlation matrix, showing within-condition correlations on the diagonal elements and between-condition correlations on the off-diagonal elements. For between-condition comparisons, elements symmetrically positioned across the diagonal (e.g. face-individual versus house-group and house-individual versus face-group) have been averaged and placed in the lower-left half of the matrix.*

As the input to a MVPA is any pattern of response, this approach can be applied across a range of measurement techniques. For example, the input pattern could be electrical field potentials across EEG electrode sites or magnetic field strengths across MEG magnetometer sites. In EEG, MVPA has only begun to be used recently, and studies more frequently include all available electrode sites, for several reasons. First, the potentials resulting from neural activity diffuse throughout the tissues in the head, such that potentials originating anywhere in the cortex may influence the readings across all electrode sites, albeit more strongly in those nearby. It is therefore impossible to select only for responses in a particular region of the brain purely by using a subset of electrode sites. This diffusion process also has the effect of spatially smoothing the patterns of response to a high degree, such that sufficient variation for correlation analyses to work may only be captured across a large spatial extent. Finally, EEG arrays often comprise far fewer recording sites across the entire head than the number of fMRI voxels available in even small ROIs; reducing this number substantially further can also impact the reliability of subsequent correlation analyses. With the added temporal element of EEG data, responses are restricted to those measured particular time-point relative to stimulus onset. Once these have been extracted, the normalisation and correlation-based MVPA can proceed in the same way as in fMRI. The analysis is then iterated across the remaining time-points in the epoch.

2.4.2. Classifier-based MVPA

Since the first implementation of MVPA using correlation-based methods by Haxby and colleagues (Haxby et al., 2001; Ishai et al., 1999), more sophisticated approaches have been developed using classification algorithms produced by the field of machine-learning. There are many different algorithms available, including support-vector machines (SVM), linear discriminant analysis and k-nearest-neighbour (Mur, Bandettini, & Kriegeskorte, 2009) and each can each be used to established whether information related to stimulus condition is encoded in patterns of neural response. This is achieved by determining whether the classifier can accurately guess the conditions of unlabelled responses after being trained on an independent set of labelled responses.

As in correlation-based methods, responses must be split into independent estimates and normalised before the MVPA can be implemented. Patterns of response are estimated across multiple independent splits of the data, with each split defining a training set and a test set. Performing multiple splits allows for the data from different trials/runs to be represented in both training and test sets in various combinations, thus helping to capture a reliable estimate of condition effects in the dataset. Classifier performance typically benefits from more training data, so the split is often configured such that the majority of responses are allocated to the training set. For example, a Leave-One-Run-Out paradigm would iterate though each sample, using it as a test case after training the classifier on all remaining responses. If there are more than two conditions, a classifier can be conducted independently across each pairwise comparison. Alternatively, each condition can be contrasted against all other conditions simultaneously by grouping them together under the same condition label.

Once the responses have been split, normalisation can then proceed in the same way as in correlation-based methods, by independently demeaning the training and test responses and, if desired, transforming to z-scores. Next, the samples constituting the training set are each projected as a point in a multi-dimensional feature-space, with each dimension/feature commonly corresponding to a voxel or electrode site. It is not recommended to input more features than samples into a classifier, so it may be necessary to reduce the dimensionality of the data at this stage. This can be done by removing lesser-modulated features, restricting the spatial extent of the ROI or by running principle components analysis of the neural activity and using a selection of the most explanatory components to define the feature space (e.g. O'Toole, Jiang, Abdi, & Haxby, 2005).

A simplified example of a classifier-based MVPA is shown in Figure 2.3. This example has just 2 features and therefore has a 2D feature-space. The training samples are derived from two different conditions, reflected by the colour of each point. The algorithm will define a decision boundary and iteratively change its parameters in order to separate the samples from the different conditions with the fewest errors possible. Decision boundaries are often linear, however, some algorithms (e.g. SVMs with radial rather than linear basis functions) make use of non-linear boundaries, which can help in delineating more complex stimulus relationships. Once the algorithm has converged (i.e. further iterations of decision boundary alterations do not improve the error rate), the generalisation of the model to the test samples can measured. The test samples are projected into the feature space – those that fall on the correct side of the decision boundary according to their condition are deemed as accurately classified and those that do not are deemed misclassified. This process is then repeated for other pairwise permutations of conditions, if necessary, and then further repeated for the different splits of the data. The accuracy of the classifier is then calculated for each contrast, condition or across the entire dataset as desired. This estimate is then compared to a baseline level of accuracy that reflects chance-level performance.

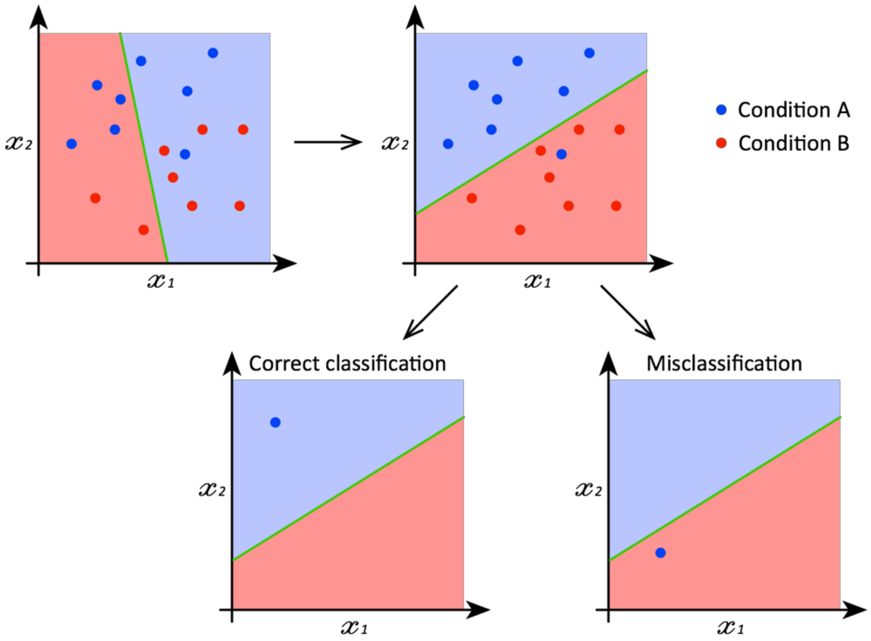

Figure 2.3 *Illustration of a simplified classification-based MVPA. In this example, responses to two conditions are measured across two electrode sites (labelled as features x_1 and x_2). Training samples for both conditions are projected into the feature space and a decision boundary is initiated with random parameters. The parameters are iteratively adjusted to separate the test samples by condition with minimal error. Once the algorithm has converged, test samples can be projected into the feature space. Samples that fall on the side of the decision boundary corresponding to its condition are deemed as a correct classification; otherwise a misclassification is declared.*

Classifier-based MVPA has the potential for greater sensitivity than correlation-based methods because a classifier can, to the extent afforded by its level of sophistication, attribute more weight to response features that more reliably distinguish the conditions while ignoring unreliable features. By contrast, voxels/electrode sites that most influence correlations between patterns are those which deviate furthest from the mean response across voxels, which do not necessarily offer the most diagnostic utility.

2.4.3. Representational Similarity Analysis (RSA)

Though correlation or classification methods are able to establish an effect of condition on patterns of response, they do not provide insight into the functional dimensions underlying the responses. To address this, a technique often referred to as representational similarity analysis (RSA; Kriegeskorte, Mur, & Bandettini, 2008; Nili et al., 2014) has been developed and widely adopted over the last decade.

An illustration of the RSA paradigm is shown in Figure 2.4. Two or more similarity matrices are generated, typically by conducting pairwise comparisons across conditions. In some cases, it may be appropriate to convert these to dissimilarity matrices, for example if one wishes to relate these matrices to analyses involving a distance measure such as hierarchical clustering or multi-dimensional scaling (Nili et al., 2014). Dissimilarity matrices can be obtained by inverting the similarity matrix, for instance by calculating one minus the correlation coefficient. In these cases, subsequent analysis may be referred to as representational dissimilarity analysis (RDA). The matrices can reflect any feature of the conditions provided that the similarity across each pairwise combination can be quantified and reduced to a single value. In addition to the similarity in patterns of neural response, matrices might contain the outcome of a behavioural study; a computational model of stimulus similarity; or a binary indicating whether conditions belong to the same higher-order group.

Once the matrices are generated, they are then compared to one another in turn, typically through correlation or regression. A high degree of similarity between two matrices suggests that the functional dimensions underlying each measure are associated. It is not recommended that within-condition values (i.e. those on the diagonal) are included in these analyses as correlation / regression coefficients between matrices can be artificially inflated if there is a strong within- versus between-condition effect (Ritchie, Bracci, & Op de Beeck, 2017). In cases where the number of conditions and consequently the number of elements in the matrix is small, a single parametric test between a group mean neural similarity matrix and a model may not be suitable. In these cases, a number of these tests may be run across individual-level neural matrices and the resulting distribution of coefficients contrasted against zero. In this way, patterns of neural response can be explicitly modelled or even contrasted across ROIs.

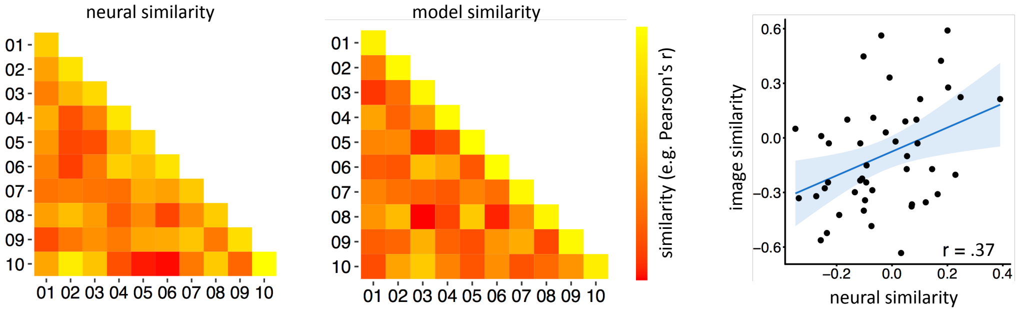

Figure 2.4 *Illustration of a representational similarity analysis (RSA) using correlation. Two similarity matrices are constructed through pairwise comparisons of each combination of conditions. The matrices can reflect any feature of the conditions provided that the similarity between conditions can be quantified and reduced to a single value. In this example, the similarity in patterns of neural response (left) and in a model (centre) have been calculated across 10 conditions. Representational similarity can then be estimated by contrasting the matrices, for instance through correlation (right). A high degree of similarity indicates a link between the functional dimensions underlying each measure.*

2.5. Defining Regions of Interest (ROIs)

It is often useful to spatially restrict fMRI analyses to a particular region of interest (ROI). The definition of an ROI allows its functional role and response properties to be investigated across a range of participants and experiments. ROIs can be defined on either an anatomical or functional basis. Anatomical ROIs are often based on physical structures of the brain, for instance the amygdala, although they can be defined by any spatial description. Functional ROIs are defined by voxels with particular response properties as measured through an fMRI paradigm, for instance showing a larger response to a particular class of stimuli. This thesis makes frequent use of ventral stream, category-selective (FFA, PPA) and retinotopic ROIs, each defined using different techniques.

2.5.1. Ventral Visual Pathway / Stream ROI

The ventral stream begins in early visual cortex and projects along the ventral surface of the occipital and temporal lobes (see Figure 2.5). Its anatomical limits are marked by the tip of the mid-fusiform sulcus; the collateral sulcus; occipito-temporal sulcus and the posterior transverse collateral sulcus (Grill-Spector & Weiner, 2014). To construct a mask of this region, we selected a series of anatomical regions of interest (ROIs) from the Harvard-Oxford cortical atlas based on these limits. 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.

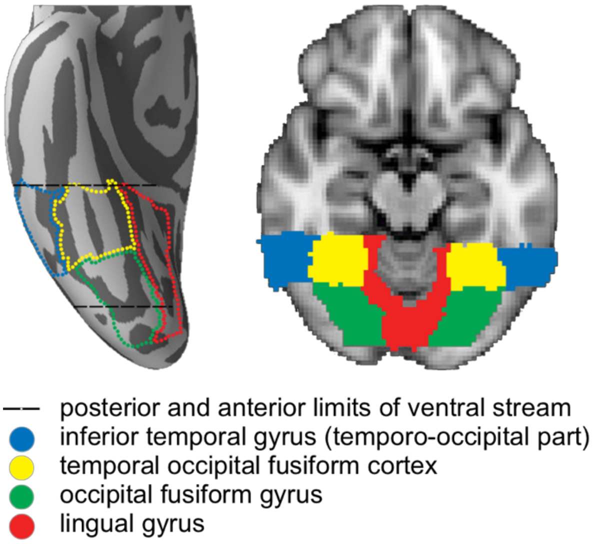
—— posterior and anterior limits of ventral stream
🔵 inferior temporal gyrus (temporo-occipital part)
🟡 temporal occipital fusiform cortex
🟢 occipital fusiform gyrus
🔴 lingual gyrus

Figure 2.5 *Ventral stream region of interest (ROI), defined anatomically. This ROI is a concatenation of four anatomical masks, shown in different colours. Posterior and anterior limits of ventral stream described Grill-Spector and Weiner (2014) is shown by dotted lines. ROI is shown on inflated surface (left) and in an axial slice of a standard-space volume (MNI-ICBM152; z = -16; right).*

2.5.2. Category-selective ROIs

FFA and PPA are regions of the ventral stream characterised by greater response to their preferred category (faces and scenes/buildings, respectively) than to non-preferred categories or scrambled images (Epstein & Kanwisher, 1998; Kanwisher et al., 1997; McCarthy et al., 1997). These functionally-defined ROIs can be revealed by measuring their categorical preferences through an independent localiser scan and a GLM of the BOLD response. In the case of FFA and PPA, these regions can be robustly highlighted through contrasting the response to face and scene stimuli, as shown in Figure 2.6, although other contrasts are often used. Once the preference for faces versus scenes has been established for each voxel, the FFA and PPA can be identified at regions of peak preference along the ventral surface of the temporal lobe, either side of the mid-fusiform sulcus. For each ROI, once a peak has been identified, a number of voxels are assigned to a mask. Determining which voxels are included can be done in several ways. Examples include selecting voxels within a given distance of the peak; with a categorical preference above a given threshold or the top *N* voxels contiguous to the peak. The lattermost is particularly useful as it allows consistent mask sizes across ROIs. Category-selective ROIs can be located independently for each subject or performed on an aggregation of the data across subjects. The former allows the analysis to be tailored to each individual's unique neural organisation. However, individual differences in combination with greater noise at the individual level makes functional localisation a more difficult and subjective process, such that ROIs may not be identified in all subjects.

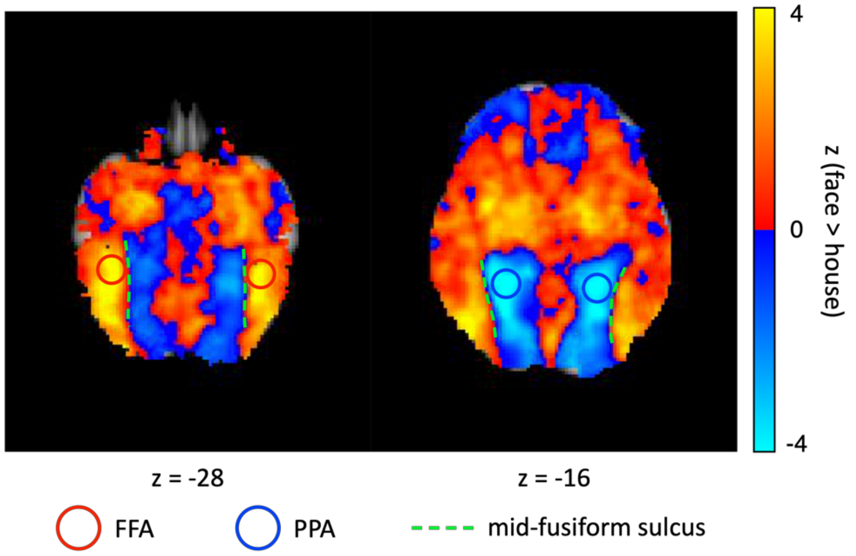

Figure 2.6 *Defining category-selective ROIs using a face versus house localiser. A GLM is used to determine the preference of each voxel in the brain to the face or house images, representing this as a z-score. Voxels with positive z-scores responded more strongly to face images and are shown in red/yellow. Voxels with negative z-scores responded more strongly to house images and are shown in blue. FFA and PPA peaks are located laterally and medially from the mid-fusiform sulcus, respectively, on the ventral surface of the temporal lobe. These locations are approximated by the red and blue rings.*

2.5.3. Retinotopic ROIs

Early visual regions are organised as retinotopic maps in which neighbouring neurons are tuned to neighbouring locations in the visual field. In order to use fMRI to measure the response properties of these regions, they first need to be localised. The maps can be revealed through additional fMRI scans in which stimuli are presented at different locations of the visual field while measuring the response in the occipital lobe. There are two popular methods for mapping the retinotopy of early visual cortex: traveling-wave and population receptive field (pRF) techniques.

2.5.3.1. Traveling-Wave Stimuli

The traveling-wave approach (Wandell et al., 2007) is designed to reveal the visual field location that each voxel in the brain is most strongly tuned to, described in terms of eccentricity (distance relative to fixation) and polar angle (direction relative to fixation). This is achieved by presenting sequences of ring and wedge shapes, as shown in Figure 2.6. In order to generate maximum neural response, the shapes are filled with a highly stimulating pattern, for instance a high-contrast, flickering checkerboard. The ring and wedge sequences are designed to measure eccentricity and polar angle representations, respectively. The ring stimulus cycles through a range of eccentricities, beginning as a small circle at fixation and expanding in a step-wise manner to become a large ring. This process is repeated a number of times such that voxels in visual field maps will respond with a particular phase. The data is analysed using Fourier analysis to determine which voxels respond with this phase. The eccentricity representation of each voxel can be determined by identifying the phase or the time point during the cycle of maximum neural activity. By taking into account the lag of the BOLD response, it is possible to associate a preferred eccentricity to each voxel. The wedge stimulus simply rotates about fixation, thus cycling through different polar angles. Each voxel’s polar angle representation is then calculated in the same way as its eccentricity.

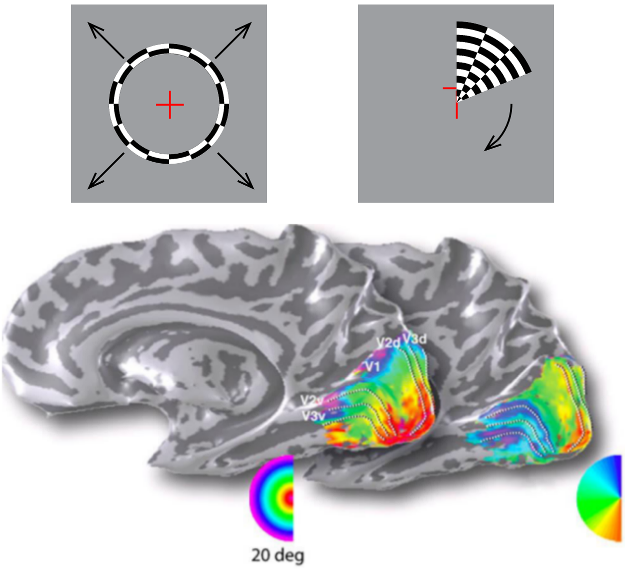

Figure 2.7 *Illustration of traveling-wave retinotopy. Stimuli are presented as flickering checkerboard patterns masked by an expanding ring (left) or a rotating wedge (right), thus cycling through different eccentricities and polar angles, respectively. These stimuli evoke a smooth travelling wave of neural activity across retinotopic regions of the visual system. Eccentricity and polar angle representations are obtained for each voxel by mapping the time-series of BOLD response onto the stimulus cycle.*

2.5.3.2. Population Receptive Field Mapping

An alternative approach to measuring visual field maps is known as population receptive field mapping (Dumoulin & Wandell, 2008). Whereas the travelling-wave approach establishes eccentricity and polar angle representations independently, pRF mapping can estimate these and other aspects of a voxel’s visual field sensitivity at once. This is achieved by modelling the area of the visual field to which the population of neurons in a voxel are responsive, known as its receptive field. Stimuli can be the aforementioned rings and wedges as well as other stimuli, for instance bar at one of four primary orientations drifting across the visual field. For each voxel, a linear model is then used to obtain estimates of the pRF coordinates and size from the time-series of the BOLD response. The bar stimulus and linear model estimation procedure are illustrated in Figure 2.8.

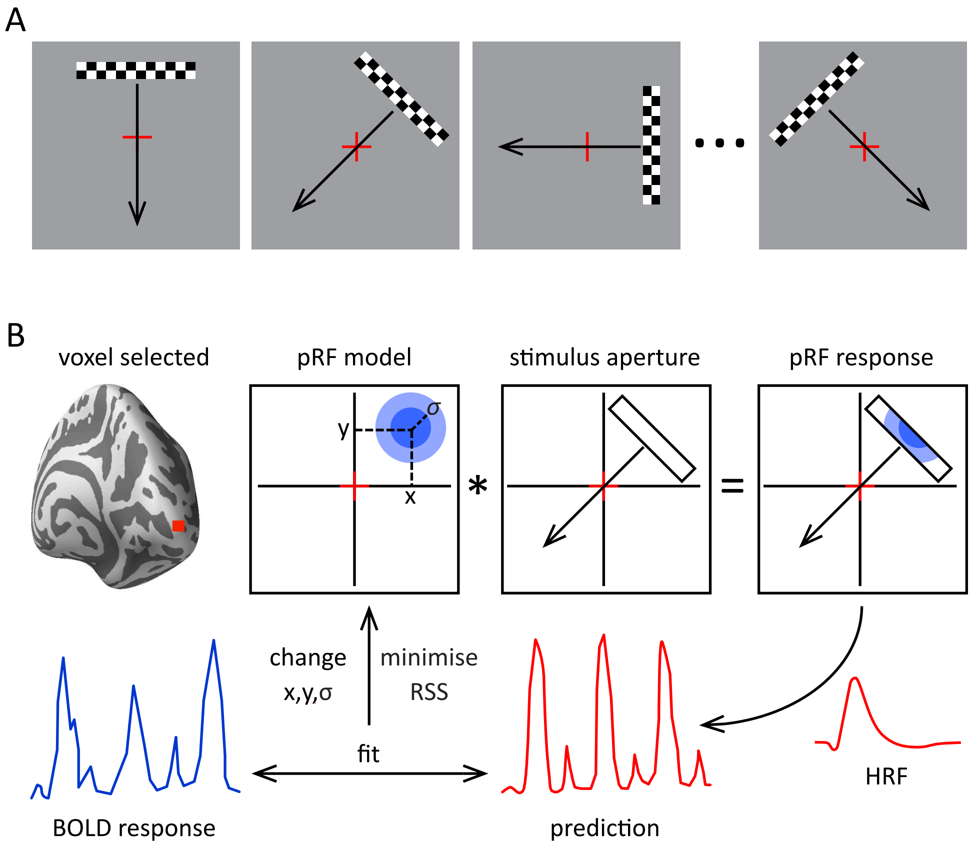

Figure 2.8 *Illustration of population receptive field (pRF) technique. (A) Drifting bar stimuli are presented at 4 primary orientations, translating smoothly across the visual field. 'Ring and wedge' type stimuli can also be used for pRF mapping. (B) Linear model for calculating coordinates and size of the receptive field for each voxel. First, the parameters are initialised as random values to generate a model of the pRF. This model is then combined with the stimulus aperture over time and convolved with the haemodynamic response function (HRF) to generate a prediction of the time-series of the BOLD response. Next, the fit between the prediction and the neural data is measured. The parameters are then iteratively adjusted, and the analysis repeated until the fit reaches a local maximum. This produces a final estimate of the pRF location and size.*

2.5.3.3. Demarcating Early Visual Regions

Once polar angle and eccentricity have been estimated for each voxel, the process of distinguishing the different early visual cortical regions can take place. Polar angle maps are particularly useful for distinguishing regions V1, V2 and V3 as the vertical meridian is represented along their borders. Transitions from one region to another can therefore be found by a reversal in the polar angle map (see Figure 2.7). Each region is divided across the 2 hemispheres, with each hemisphere representing the contralateral visual hemifield. For V2 and V3, these regions are then further divided into ventral and dorsal aspects representing the upper and lower quarter-fields, respectively.

2.6. Image Properties

The aim of this thesis is to explore the extent to which patterns of response in high-level visual cortex can be explained by variance in the image properties of objects. In order to understand the impact of image properties, it is important to be able to measure these properties. This thesis uses the GIST descriptor (Oliva & Torralba, 2001) – an image analysis tool that attempts to capture the key spatial properties of an image. It measures the spectral properties (spatial frequency and orientation) at different locations across the spatial extent of the image, condensing this information into a vector. This process is shown in Figure 2.9. GIST descriptions can be calculated for a number of different images and then contrasted with one another, for instance using correlation. Although originally developed for scene perception in computer vision, the GIST descriptor offers a neurologically plausible model of the low-level image properties represented in the visual system. It directly measures spatial and spectral properties known to be strongly encoded by early visual cortex (Hubel & Wiesel, 1968; Wandell et al., 2007) and to some extent by high-level regions (Arcaro et al., 2009; Nasr et al., 2014; Rajimehr et al., 2011), and has proved useful in predicting patterns of ventral responses to different scenes (Watson et al., 2014) and objects (Andrews et al., 2015; Rice et al., 2014).

Although we chose to use the GIST descriptor, it should be noted that there are other statistical image descriptors available. Examples include the HMAX model (Riesenhuber & Poggio, 1999) and the SIFT descriptor (Lowe, 2004), while recent advances in deep learning have also produced successful models (Szegedy et al., 2015). It is possible that some of these alternatives offer more neurologically plausible and / or accurate predictions of neural responses. However, a full comparison of these tools is beyond the scope of this thesis. Instead, the aim is to identify instances where the GIST descriptor, as a model of low-level image properties, can successfully predict neural responses. In these instances, the image properties captured by GIST can be said to play a significant role in driving the neural responses.

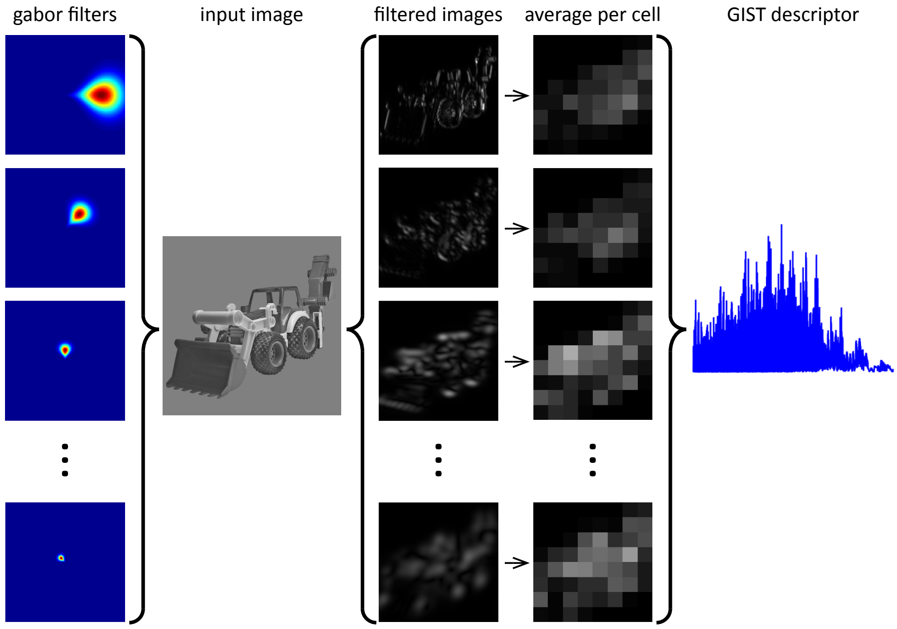

Figure 2.9 *Illustration of the GIST descriptor applied to an example image. First a number of filters are generated across a predetermined number of spatial frequency and orientation bandwidths (in this example 4 and 8, respectively). Each of these filters is then applied independently to the image to produce a number of filtered images. The spatial resolution of the filtered images is then downsampled to a predetermined size (in this case 8x8), giving an average pixel intensity in each cell. Finally, the downsampled images are reshaped into a single vector, with each value representing the pixel intensity in one of the cells in one of the images. This results in a vector that describes the spatial frequencies and orientations present across the image space*

Chapter 3. The Emergence of an Image-Based Representation of Objects in High-Level Visual Areas

This chapter is adapted from: Coggan, D. D., Watson, D. M., Hartley, T., Baker, D. H., & Andrews, T. J. (under review). The Emergence of an Image-Based Representation of Objects in High-Level Visual Areas. *Cerebral Cortex*.

3.1. Abstract

We encounter a vast number of images during a life-time of natural viewing. However, our ability to understand how objects are represented in the human brain is limited by the fact that only a finite number of images can be shown during a typical experiment. This could lead to an uneven sampling of the image space that biases our understanding of the way objects are represented. To address this issue, we developed a novel data-driven approach to stimulus selection in which a large database of objects was described in terms of their image features (orientation, spatial frequency, spatial position). A clustering algorithm was used to evenly select clusters of objects from regions of this feature space. Our aim was to determine, using fMRI, how these object clusters were represented along the hierarchy from low-level to high-level visual regions. Although the clusters did not correspond to typical object categories, they elicited distinct patterns of response in both low-level and high-level visual cortex. However, the representation of the object clusters changed from low-level to high-level regions. Responses in high-level

ventral regions corresponded more with the perceptual similarity of the object clusters. This suggests the emergence of an image-based neural representation in high-level visual cortex that is important for object perception.

3.2. Introduction

Patterns of neural response in higher-visual areas of the ventral visual pathway have been linked to higher-level properties of objects (Chao, Haxby, & Martin, 1999; Connolly et al., 2012; Haxby et al., 2001; Konkle & Oliva, 2012; Kriegeskorte, Mur, Ruff, et al., 2008; Naselaris et al., 2009). 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 et al., 2015). For example, a number of studies have shown that low-level differences in the visual properties of objects can explain a significant amount of the variance in category-selective regions of visual cortex (Bracci & Op de Beeck, 2016; Proklova, Kaiser, & Peelen, 2016; Rice et al., 2014; Watson et al., 2014; Watson, Young, et al., 2016). Moreover, category-selective patterns of response are still evident when images have been scrambled in a way that preserves their visual properties, but removes their semantic properties (Watson, Andrews, & Hartley, 2017).

Despite this progress, a fundamental constraint in our ability to explore how objects are represented in the brain is that only a finite number of images can be presented in a typical neuroimaging experiment. This has led to experimental designs that contrast responses to stimuli from experimenter-defined categories, which makes it difficult to disentangle the subjective manipulation of higher-level dimensions of the stimulus from those driven by correlated lower-level dimensions. Many studies have made considerable efforts to control for such confounds by directly comparing the influence of low-level and high-level properties on patterns of response (Bracci & Op de Beeck, 2016; Clarke & Tyler, 2014; Proklova et al., 2016). Nevertheless, these studies do not overcome the limitations posed by the subjective sampling of the available stimulus space. To understand how the neural representation of objects might emerge, it is necessary to develop methods to sample objects in an objective 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. Images from a large object database were described in terms of their image properties and clustering algorithms were used to evenly sample distinct clusters of objects from this image space. The logic is that these object clusters will provide a good approximation to the diversity of objects that an individual will be exposed to during a life-time of natural viewing. Our aim was to determine how these object clusters are represented from low-level to high-level visual areas using fMRI. We were particularly interested in how the representations changed from low-level to high-level regions. We complemented these data with an analysis of the perceptual similarity of the object clusters. This allowed us to compare our perception of the object clusters with the corresponding neural representation at different stages of processing. The results show that the representation of the data-driven object clusters changes from low-level to high-level regions. Moreover, the representation in high-level regions of the ventral stream is better predicted by perception. Overall, these results show how an image-based representation of objects that is selective for the perceptual properties of objects could emerge in high-level regions of the ventral visual pathway.

3.3. Methods

3.3.1. Data-driven Image Selection

The experimental stimulus set was generated by an entirely data-driven approach. 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). Image properties were 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. We configured the descriptor to measure the energy at 8 spatial frequencies across 8 orientations and 64 spatial subdivisions (8x8) of the image, resulting in a vector of 4096 values that described each image.

Images were first cropped and resized to the resolution at which they would be presented in the experiment (720x720 pixels) and converted to greyscale. A GIST descriptor was then generated for each image. GIST vectors were next normalised by first scaling each component of the vectors to sum to 1 across images, and second by scaling each vector to have a magnitude of 1. Each image is thus represented as a point in a 4096-dimensional feature space by its normalised GIST descriptor. Attempting to apply clustering algorithms in such a high-dimensional space can be problematic, so we first reduced the dimensionality using principal components analysis (PCA). The first 20 principal components were selected; these explained 58.04% of the variance of the original components. We applied a k-Means clustering algorithm (k = 10; Euclidean distance metric) to identify 10 distinct clusters of images within this space, such that images within a cluster are defined by having similar visual properties to one another. Finally, we selected the 24 images nearest the centroid of each cluster as measured by

Euclidean distance. This process is illustrated in Figure 3.1A. The GIST descriptor is not sensitive to colour, so images were presented in greyscale. PCA and k-Means algorithms were implemented using the Python Scikit-learn toolbox (Pedregosa et al., 2011). Following this, a correlations similarity matrix was constructed by correlating the principal component vectors within and between conditions using a leave-one-image-out cross-validation procedure (Figure 3.1B). 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 3.1C). Two versions of this stimulus set were then created by applying a uniform, mid-grey background (untextured condition, Figure 3.1D) and a unique, pink noise (1/f) background (textured condition, Figure 3.1E) to each of the 240 images. The entire object stimulus set prior to the addition of backgrounds is show in Supplementary Figure A.1.1.

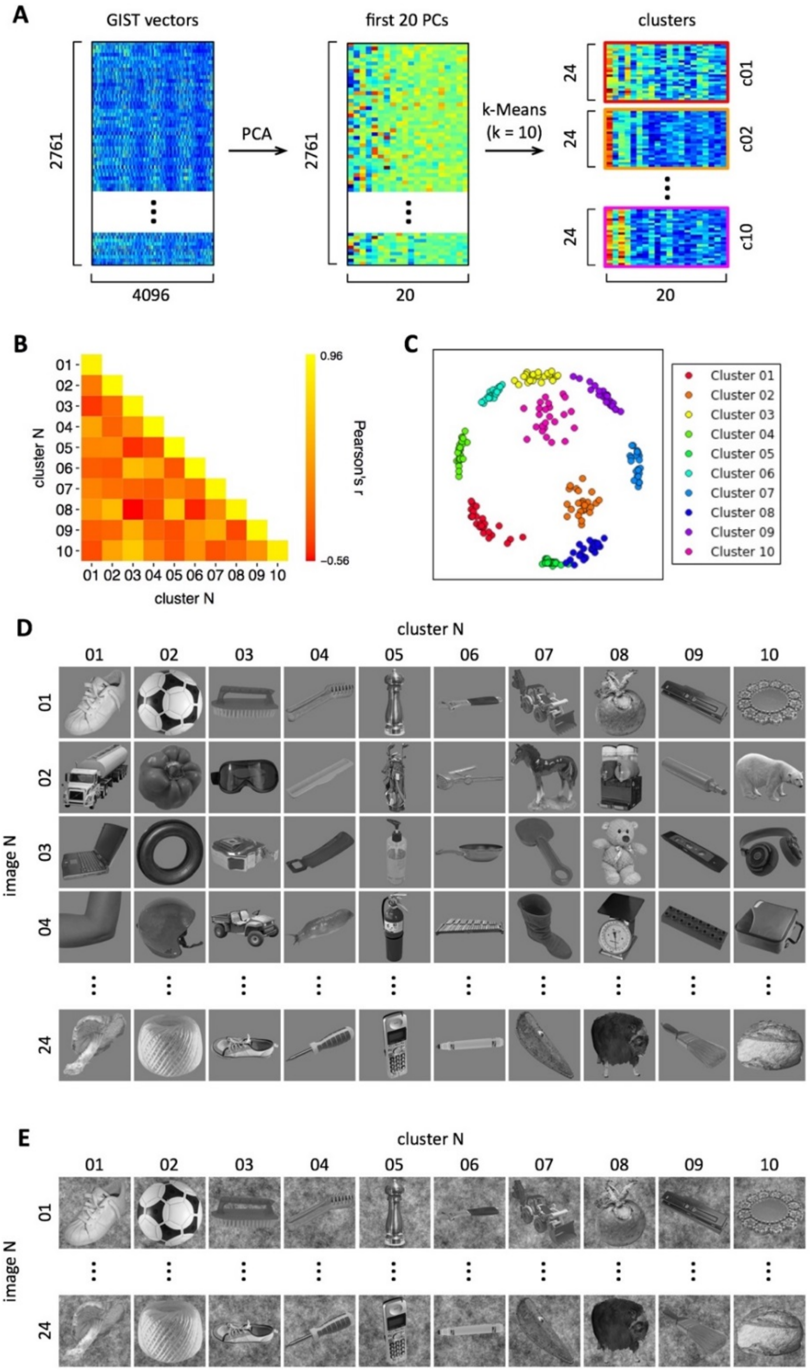

Figure 3.1 *Data-driven image selection. (A) GIST descriptions were generated for each image in the BOSS database. PCA was used to reduce the dimensionality of the data, with the first 20 PCs selected. 10 distinct clusters within this feature space were then defined through k-means clustering, with the 24 most proximate images to each cluster centroid selected to represent different conditions. (B) Correlation matrix showing similarity in GIST descriptions within and between the different conditions. (C) Multidimensional scaling plot approximating the locations of the selected images within the feature space. (D) Examples of stimuli from each of the 10 clusters on a uniform, mid-grey background ('untextured' condition). (E) The same stimulus set was also super-imposed on a pink noise background ('textured' condition).*

3.3.2. fMRI experiment

3.3.2.1. Participants

Twenty-one participants took part in the fMRI experiment (11 male, mean age = 23.0, SD = 2.7 years). Sample size was based on previous studies using similar designs (Coggan et al., 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 Declaration of Helsinki.

3.3.2.2. Design and Procedure

The fMRI experiment consisted of two scans, each lasting 10 minutes. 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, 6 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. Each object was shown once per scan, with each cluster represented across four blocks. The order of the blocks was randomized for each subject and scan. Participants performed a task while 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 placed on 40 of the 240 images presented throughout the scan, selected at random for each subject. 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, with all images subtending approximately a 15° retinal angle. The image surround consisted of the same mid-grey shade as the background for untextured images, such that it blended seamlessly with the border of images from this condition. Stimulus presentation was controlled through Psychopy (Peirce, 2007).

3.3.2.3. Data Acquisition

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. 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, voxel dimensions = 2.25 x 2.25 x 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 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 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 = 9.98ms, FOV = 288 × 288 mm, matrix size = 512 × 512, voxel dimensions = 0.56 × 0.56 x 3 mm, flip angle = 90°), then to a high-resolution T1-anatomical image (TR = 7.96ms, TE = 3.05ms, FOV = 290 × 290 mm, matrix size = 256 × 256, voxel dimensions = 1.13 × 1.13 x 1 mm, flip angle = 20°) and finally onto the standard MNI brain (MNI-ICBM152).

3.3.2.4. Regions of Interest (ROIs)

We used a ventral stream ROI (described in section 2.5.1) with voxels overlapping with early visual cortical regions (V1,V2,V3,hV4) removed. Masks of early visual areas were based on probabilistic visual field maps developed by Wang and colleagues (Wang, Mruczek, Arcaro, & Kastner, 2015). These maps were, however, used in a separate analysis. Our rationale for using these masks was to determine how the representation of objects changes from early to higher levels of visual system. A few regions (MST/TO2, SPL1, IPS4, IPS5, FEF) were excluded from our analysis as they contained too few voxels (<22, once converted to 2mm MNI space and restricted to the field of view). All ROIs are shown in Figure 3.2.

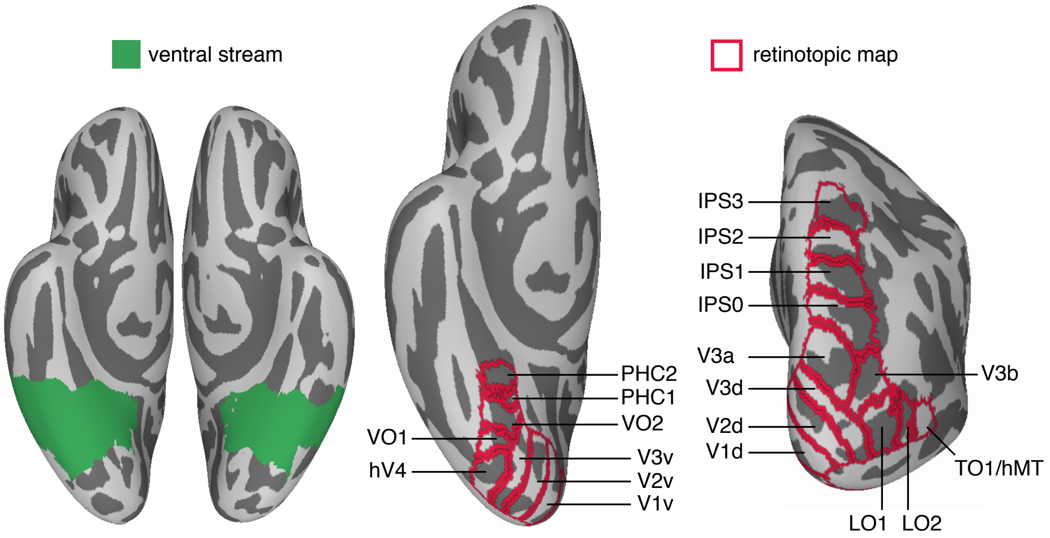

Figure 3.2 *Regions of interest, projected onto inflated cortex. Only right hemisphere aspects of bilateral retinotopic ROIs are shown.*

3.3.2.5. Multi-voxel Pattern Analysis

The reliability of patterns of neural response to each object cluster was tested using a leave-one-participant-out (LOPO) cross-validation paradigm (Poldrack, Halchenko, & Hanson, 2009; Rice et al., 2014). Parameter estimates were normalised by subtracting the mean response per voxel per subject 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.).

3.3.3. Perceptual similarity experiment

To generate a model of the similarity between object clusters based on human perceptual judgements, we recruited twenty participants (11 male, mean age = 35.7, SD = 17.2 years) for a behavioural study involving a card-sorting task (Jenkins, White, Van Montfort, & Burton, 2011). None of the participants had taken part in the fMRI experiment, and all had normal or corrected-to-normal vision and no history of mental illness. Informed, written consent was obtained for all participants and the study was approved by the York Psychology Department Ethics Committee.

Each participant was provided with a set of printed cards measuring 4 x 4 cm that comprised a subset of the images (60 images; 6 per cluster). Subsets were counterbalanced across participants. Participants were required to sort the cards into 10 or fewer piles according to their perceptual similarity, such that cards within a stack were ones that they perceived to all be similar to one another. The task was designed to allow participants as much freedom as possible to sort the cards however they wished. The precise definition of perceptual similarity was left deliberately vague so as to encourage participants to form their own interpretation. Card piles were allowed to vary in size, and participants were allowed unlimited time to complete the task. In order to prevent the paradigm becoming a memory task, participants were required to stack cards next to one another so that they could always be seen. Participants were then asked to provide a label for each pile that reflected the perceptual property (or properties) common across images within the pile.

Following the test, the number of cards from each of the object clusters was counted for each of the card stacks. A vector was constructed for each cluster representing the counts for that cluster across each of the card stacks (Figure 3.3A). The lower-triangle of a perceptual similarity matrix was constructed by taking the dot-product of the vectors between each unique pairing of clusters, such that the element at position (*i,j*) represents the dot product between the vectors of the *i*th and *j*th scene clusters respectively (Figure 3.3B). Values thus represent the frequency of co-occurrence of images from different object clusters across card stacks. Finally, to see which perceptual properties subjects used to group objects, all stack labels were collated and entered in Wordle (www.wordle.net), which generates a graphical depiction of the frequency of each word across a text document (Figure 3.3C).

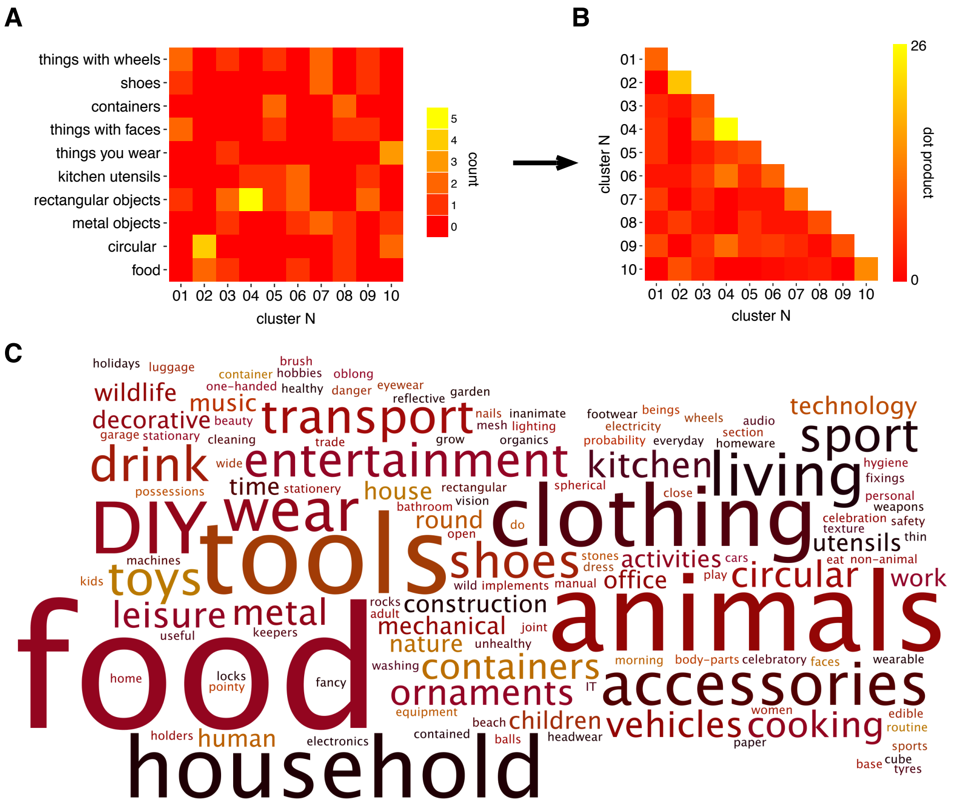

Figure 3.3 *Analysis of perceptual card-sorting task.* ***A*** *Grouping of stimuli from different clusters into perceptually similar stacks by an example subject.* ***B*** *Matrix showing, for each pair of clusters, the dot product between their vectors from matrix* ***A****.* ***C*** *Wordle analysis of stack labels provided by subjects in the perceptual task. Font size directly corresponds to frequency of occurrence across the dataset, with more frequent words shown at larger sizes.*

3.4. Results

Figure 3.4A shows the average pattern of response to each of the object clusters in the ventral visual pathway ROI. Figure 3.4B shows the similarity in patterns of response within and between different clusters. 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). 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.

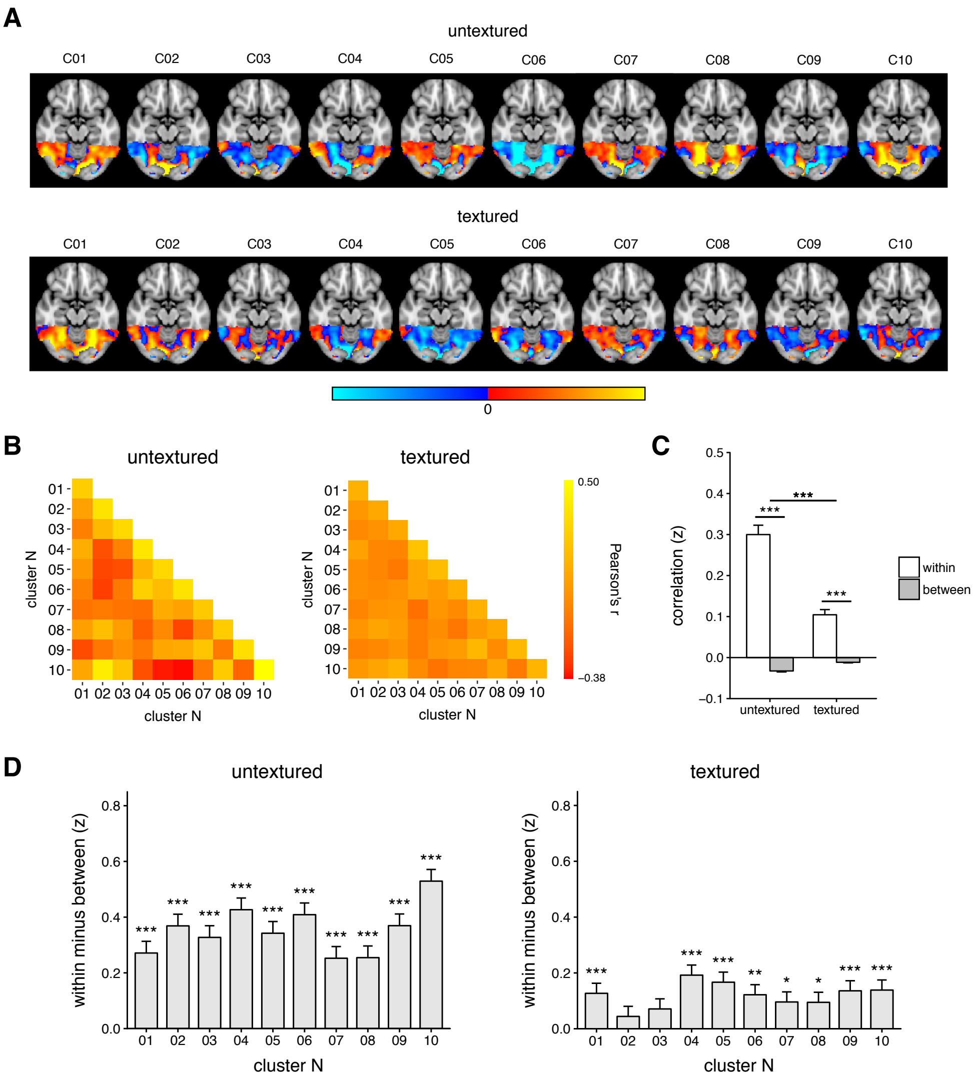

Figure 3.4 *MVPA.* ***A*** *Patterns of neural response in the ventral visual pathway to different object clusters with untextured and textured backgrounds. Patterns of response were normalised for each background type by subtracting the voxel-wise mean response across all 10 clusters from the response to each cluster. Red and blue colour maps thus indicate values above and below the mean response respectively.* ***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 = 0.83, p<0.001; between-cluster correlations only).* ***C*** *Bar plot showing within-cluster and 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 standard error of the mean. * p < .05, ** p < .01, *** p < .001*

We found that different object clusters evoked distinct patterns of neural response across the ventral visual ROI. A 3-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(1,29) = 157.88, $\eta_G^2 = 0.45$, p < .001), with within-cluster correlations being higher than between-cluster correlations. However, there was an interaction with background (F(1,20) = 91.20, $\eta_G^2 = 0.17$, p < .001), suggesting that the distinctiveness of cluster-specific patterns differed across background types (Figure 3.4C). Post hoc analysis revealed higher within-cluster than between-cluster correlations for both untextured (t(20) = 13.16, Cohen's d = 2.87, p < .001) and textured backgrounds (t(20) = 8.25, Cohen's d = 1.80, p < .001), but a stronger effect for untextured images (t(20) = 9.75, Cohen's d = 2.13, p < .001). There was also a two-way interaction between comparison and cluster (F(9,180) = 3.75, $\eta_G^2 = 0.04$, p < .001) and a three-way interaction between background, cluster, and comparison that approached significance (F(9,180) = 1.87, $\eta_G^2 = 0.02$, p = .060). To investigate these effects, post-hoc pairwise comparisons of the within-cluster and between-cluster correlations were determined for each background-cluster combination (Figure 3.4D). For the untextured background, there were significantly higher within-cluster than between-cluster correlations for all clusters (t(20) > 5.45, Cohen's d > 1.79, p < .001). For the textured background, eight of the ten clusters showed higher within- than between-cluster correlations (t(20) > 2.55, Cohen's d > 0.85, p < .039). The non-significant clusters were clusters 2 (t(20) = 1.06, Cohen's d = 0.36, p = .234) and 3 (t(20) = 1.68, Cohen's d = 0.56, p = .112).

Our next analysis investigated the extent to which the pattern of neural response in the ventral visual pathway could be predicted by the perceptual and visual properties of the image clusters. To measure perceptual similarity, participants completed a card sorting task in which they had to sort images from the different object clusters into piles based on perceptual similarity of the images (see Figure 3.3). First, we asked whether images from the same cluster were more likely to be sorted together than images from different clusters. Using the same within-cluster versus between-cluster approach used for patterns of neural response, we found the images from the same cluster were more likely to be in the same pile in the perceptual task (t(9) = 12.97, Cohen's d = 4.10, p <.001). At the end of the sorting task, participants were asked to label the piles. We used a Wordle analysis (wordle.net) to explore the terms people used. Despite the fact that participants were instructed to sort the images based on their perceptual similarity, the labels used often reflected higher-level descriptions (Figure 3.3C).

Next, we asked whether the pattern of perceptual sorting predicted the patterns of neural response. For example, if images from two clusters are perceived to be similar to each other, are the patterns of neural response also similar? To prevent differences in within- and between-cluster correlations artificially inflating correlations between matrices, our analysis was only performed on the between-cluster comparisons. There was a positive correlation between the neural correlation matrix from the ventral visual pathway and the perceptual similarity matrix (untextured: r = .44, p = .002; textured: r = .32, p = .032), suggesting that clusters eliciting similar patterns of response were perceived as similar by human observers (Figure 3.5A). We then asked whether the patterns of neural response could be explained by the similarity in image properties derived from the GIST descriptor. Again, there was a significant, positive correlation between the neural correlation matrix and the image (untextured: r = .37, p = .011; textured: r = .38, p = .010), suggesting that clusters with more similar image properties were also likely to elicit more similar patterns of neural response (Figure 3.5B).

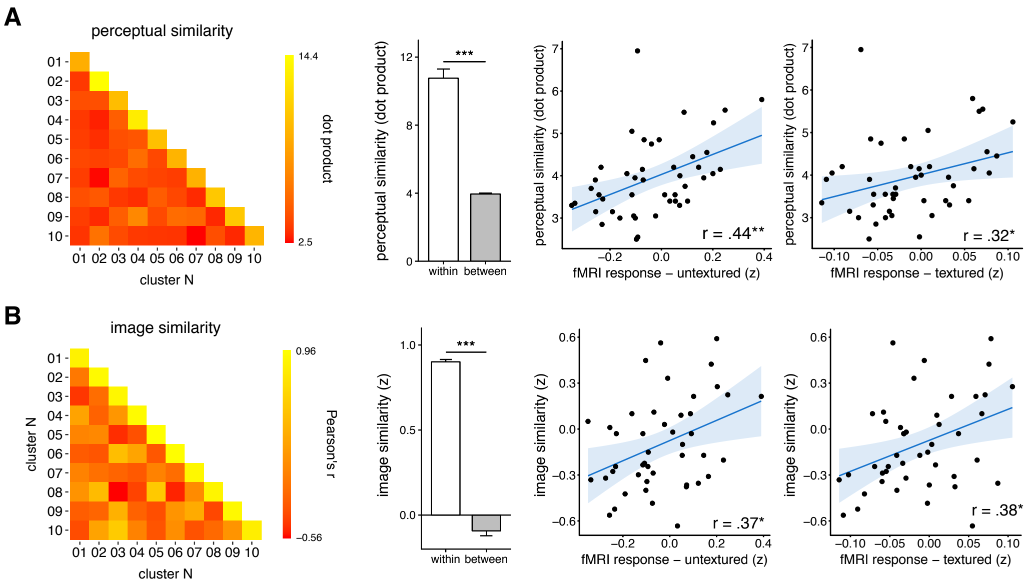

Figure 3.5 Analysis of perceptual (**A**) and image (**B**) models. Matrices show similarity based on human perceptual judgements and GIST descriptions. Bar plots show within- and between-cluster values for either model. Error bars represent standard error of the mean. Scatterplots show correlation between models and the neural matrices for untextured and textured images. 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

To explore how the neural representation of the object clusters changed along the visual processing hierarchy, we used probabilistic visual field map ROIs (Wang et al., 2015). These maps include early visual areas in the posterior occipital lobe, as well as ventral and dorsal stream areas in the temporal and parietal lobes. All areas in the posterior occipital lobe showed higher within-cluster compared to between-cluster values demonstrating that there were distinct patterns of response. Similarly, in occipito-temporal cortex, all regions on the ventral surface and 4 out of 5 regions on the lateral surface showed distinct patterns of response to the different object clusters. However, only 1 of the 4 regions on the dorsal surface showed a distinct pattern to different clusters. Test statistics for each ROI are shown in Table 3.1.

location	region	t(20)	Cohen's d
posterior	V1d	8.62***	1.88
	V1v	9.90***	2.16
	V2d	8.40***	1.83
	V2v	11.58***	2.53
	V3d	7.74***	1.69
	V3v	10.58***	2.31
	hV4	7.93***	1.73
ventral	VO1	5.46***	1.19
	VO2	5.53***	1.21
	PHC1	3.84**	0.84
	PHC2	4.68***	1.02
lateral	V3a	4.49***	0.98
	V3b	5.23***	1.14
	LO1	4.98***	1.09
	LO2	1.36	0.30
	hMT/TO1	5.62***	1.22
dorsal	IPS0	1.00	0.21
	IPS1	2.32*	0.51
	IPS2	-0.83	0.18
	IPS3	-0.33	0.07

Table 3.1 *Results of t-tests comparing within-cluster and between-cluster correlations for retinotopic ROIs. Positive t values reflect higher values for within-cluster correlations.*

** p < .05, ** p < .01, *** p < .001*

To determine how the neural representation of objects changes across the visual hierarchy, we compared the neural correlation matrices across these different regions (Figure 3.6A). For each region, between-cluster correlations from the z-transformed correlation matrices were selected and compared to each of the other 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 (Figure 3.6B). 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.

To determine whether this difference between low-level and high-level regions was linked to the perception of objects, we regressed the untextured neural similarity matrix for each subject in each region with the perceptual similarity matrix (Figure 3.5A) and image similarity matrix (Figure 3.5B). The perceptual and image matrices were scaled between zero and one prior to the regression and only off-diagonal elements of the matrices were used. Neural and image matrices were z-transformed prior to the analysis. Perceptual and image matrices were entered simultaneously into a multiple regression with the neural matrix from each subject. This analysis produced a beta value for each combination of subject, model (perceptual, image) and ROI. These values were tested against zero using one-sampled t-tests. This analysis was first performed for the ventral visual pathway ROI. The beta values for both perceptual (mean = 0.29, SEM = 0.03, t(20) = 9.63, Cohen's d = 2.10, p < .001) and image (mean = 0.17, SEM = 0.02, t(20) = 10.19, Cohen's d = 2.22, p < .001) models were significant. Moreover, the values for the perceptual matrix were significantly higher than those for image matrix (t(20) = 3.59, Cohen's d = 0.78, p < .001). The analysis was then run on each of the visual field ROIs (Figure 3.6CD). Statistics for each comparison across all the ROIs are shown in Table 3.2. These show a gradual decline in the importance of image properties from posterior regions. However, there is a general increase in the importance of perceptual properties in the ventral visual field regions.

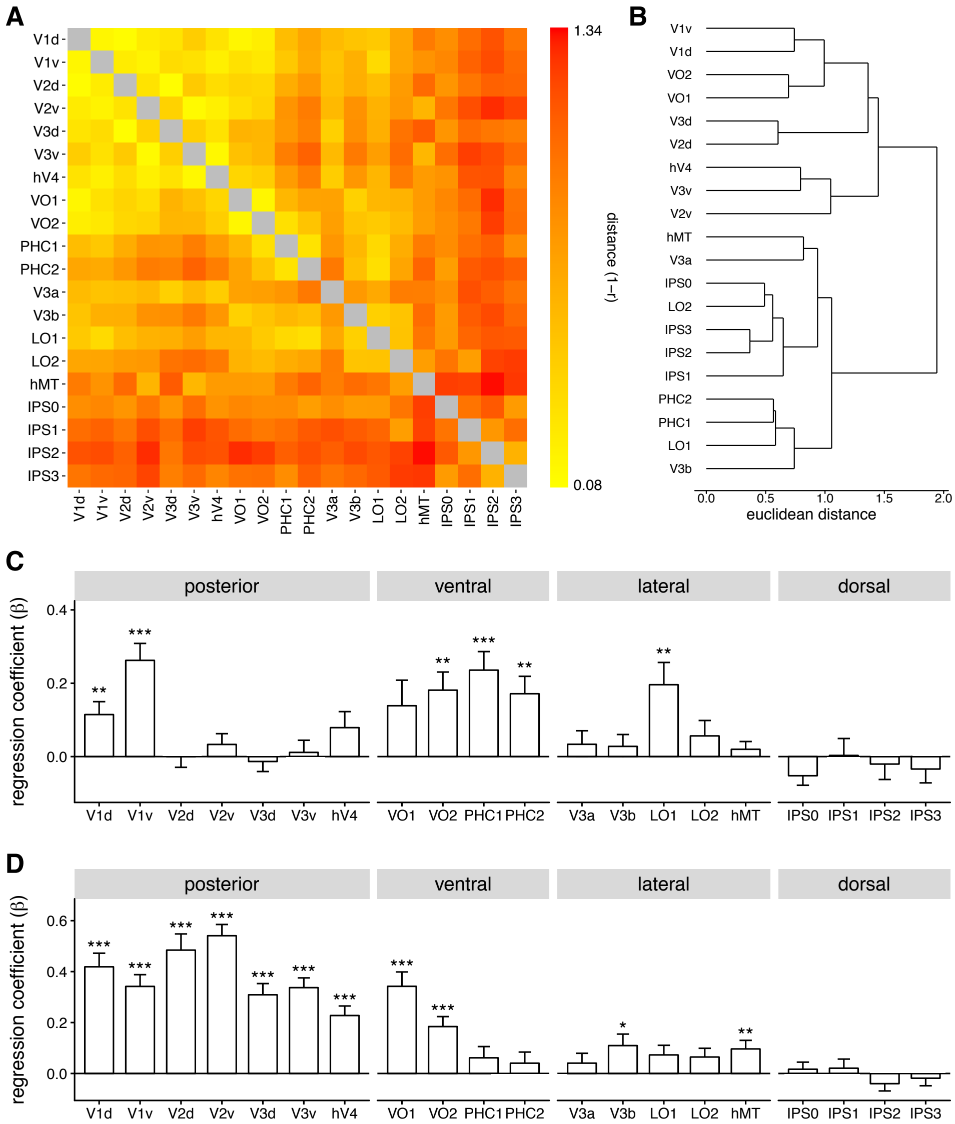

Figure 3.6 ROI analyses. **A** Representational similarity of retinotopic regions based on the neural similarity matrices in the untextured condition. **B** Hierarchical clustering of regions based on Euclidean distance. Beta coefficients for perceptual (**C**) and image (**D**) matrices when regressed against untextured neural similarity matrices at each ROI. * p < .05, ** p < .01, *** p < .001 (tested against 0).

		perceptual		image		perceptual vs image	
location	region	t(20)	Cohen's d	t(20)	Cohen's d	t(20)	Cohen's d
posterior	V1d	3.25**	0.71	7.83***	1.71	-5.90***	1.29
	V1v	5.68***	1.23	7.42***	1.62	-1.25	0.27
	V2d	-0.01	0.00	7.63***	1.67	-7.11***	1.55
	V2v	1.12	0.25	12.24***	2.67	-9.13***	1.99
	V3d	-0.49	0.11	7.00***	1.53	-5.47***	1.19
	V3v	0.34	0.08	8.78***	1.92	-6.67***	1.45
	hV4	1.81	0.39	6.07***	1.33	-2.39*	0.52
ventral	VO1	1.99	0.43	6.06***	1.32	-2.47*	0.54
	VO2	3.67**	0.80	4.69***	1.02	-0.05	0.01
	PHC1	4.66***	1.02	1.39	0.30	2.57*	0.56
	PHC2	3.61**	0.79	0.93	0.20	1.83	0.40
lateral	V3a	0.91	0.20	1.05	0.23	-0.11	0.03
	V3b	0.86	0.19	2.42*	0.53	-1.39	0.30
	LO1	3.23**	0.71	1.93	0.42	1.38	0.30
	LO2	1.34	0.29	1.89	0.41	-0.13	0.03
	hMT/TO1	0.94	0.21	2.86**	0.62	-1.76	0.38
dorsal	IPS0	-2.01	0.44	0.60	0.13	-1.56	0.34
	IPS1	0.07	0.01	0.58	0.13	-0.29	0.06
	IPS2	-0.48	0.11	-1.40	0.30	0.41	0.09
	IPS3	-0.90	0.20	-0.63	0.14	-0.30	0.06

Table 3.2 *Results of t-tests comparing beta values against zero for multiple regression of untextured neural matrices onto perceptual and image similarity matrices for each ROI.*

** p < .05, ** p < .01, *** p < .001*

3.5. Discussion

The aim of this study was to compare how objects are represented along the ventral visual pathway. A key feature of our approach is the use of data-driven methods for image selection. A limitation of many neuroimaging studies is that only a finite number of images can be shown during a typical experiment. This contrasts with the vast number of images that a typical person encounters during a life-time of natural viewing. Thus, stimuli selected in neuroimaging experiments may not sample image space in a uniform way, making it difficult to separate the effects of arbitrary and subjective manipulations of stimulus conditions from those driven by more basic dimensions. In this experiment, a data-driven approach was used to sample objects based on the distribution of image properties found across a large number of natural objects. Our aim was to explore how these objects were represented along the visual hierarchy.

Images selected from different regions of this natural image space gave rise to distinct patterns of neural response throughout visual cortex. The ability of low-level visual areas to discriminate object clusters that differ systematically in their image properties is not surprising given the topographic maps found in these regions, which are tightly linked to the properties of the visual image (Hubel and Wiesel, 1968; Wandell et al., 2007). However, the organization of high-level regions is thought to be based on the conceptual or semantic properties of objects (Kanwisher, 2010; Grill-Spector & Weiner, 2014; Connolly et al., 2012; Haxby et al., 2001; Kriegeskorte et al., 2008; Naselaris et al., 2009; Konkle & Oliva, 2012). It has proved difficult to explain how selectivity for object categories suddenly emerges from these low-level representations (Op de Beeck et al., 2008). The distinct patterns of response we observe suggests that the ventral visual pathway is also sensitive to image properties. This finding is consistent with previous studies showing that image properties predict patterns of response to objects in the ventral visual pathway (Rice et al., 2014; Andrews et al., 2015; 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 reinforces the importance of image properties in the neural representation of this region.

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, the data clearly shows that the neural representation changes along the visual hierarchy (see Figure 3.6). 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 (Sigman, Cecchi, Gilbert, & Magnasco, 2001; Geisler, 2008; Oppenheim & Lim, 1981; Thomson, 1999). Our data suggest that high-level regions represent combinations of image properties typically found in natural objects, whereas low-level regions have a more homogeneous representation of image properties.

To determine how these patterns of neural response across visual cortex predicted the perceptual properties of the object clusters, participants performed a sorting task with the images (Jenkins et al., 2011; Watson et al., 2017). In the perceptual task, participants grouped images of objects based on their perceptual similarity. Despite the fact that images from each object cluster did not appear to belong to particular categories, objects from the same cluster were perceived to be more similar than objects from different clusters. Interestingly, it would appear that participants were not necessarily aware that they were sorting based on the image properties, because they predominantly labelled the stacks according to higher-level properties of the images. Nevertheless, the pattern of neural response to object clusters in high-level regions of the ventral stream was predicted by the perceptual similarity of the objects. We found that the perception of the object clusters predicted the pattern of neural response in these high-level ventral regions more than the image properties. This is significant as it shows a direct link between patterns of neural response in the ventral stream and the perception of these object clusters and suggests that the image-based representation in high-level regions is tuned to the key properties necessary for the discrimination of natural images. In contrast, patterns of response in low-level regions (V1-V4) were predicted more by the image properties of the clusters.

An obvious advantage of a relatively image-based representation in high-level visual cortex is that it can be used more flexibly in the 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, Mur, Ruff, et al., 2008; Naselaris et al., 2009), animacy (Chao et al., 1999; Kriegeskorte, Mur, Ruff, 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. This would not be inconsistent with the distinct patterns of response that are evident to higher-level properties of objects. However, a common image-based representation would allow for the extraction of different information depending on the task.

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. 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 untextured condition were presented within a square of pink noise, reducing the salience of this diagnostic cue. 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, Young, et al., 2016). The spatial envelope is unusually clear in images of isolated objects, and it might be argued that the distinctiveness of cluster-specific responses is specific to these non-naturalistic stimuli. 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 generalize to natural images in which the ability to separate figure and ground is likely to be an important processing step (Rubin, 2001).

An important feature of our findings is that the spatial patterns of response to different object clusters generalized 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 organization of these regions. In many previous MVPA studies, the analysis 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 (Rice et al., 2014; Flack et al., 2015; 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 organizing 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 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 and contributes to perceptual judgements.

Chapter 4. Category-Selective Patterns of Neural Response in the Ventral Visual Pathway in the Absence of Categorical Information

This chapter is adapted from: 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.

4.1. 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.

4.2. Introduction

In the previous chapter, we showed that different clusters of low-level properties evoke distinct patterns of neural response in the ventral visual pathway. This is in line with the notion that a tuning to more basic visual properties might underlie category-selective patterns of neural response across the ventral visual pathway. However, due to the correlation between low-level and high-level properties of the image, it is possible that the observed effects may be due to semantic differences between clusters. In other words, just as a low-level tuning might underlie category effects, a categorical tuning might underlie low-level effects. A more direct contrast between the roles of low- and high-level visual properties is therefore warranted. 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.

4.3. Methods

4.3.1. 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 Figure 4.1.

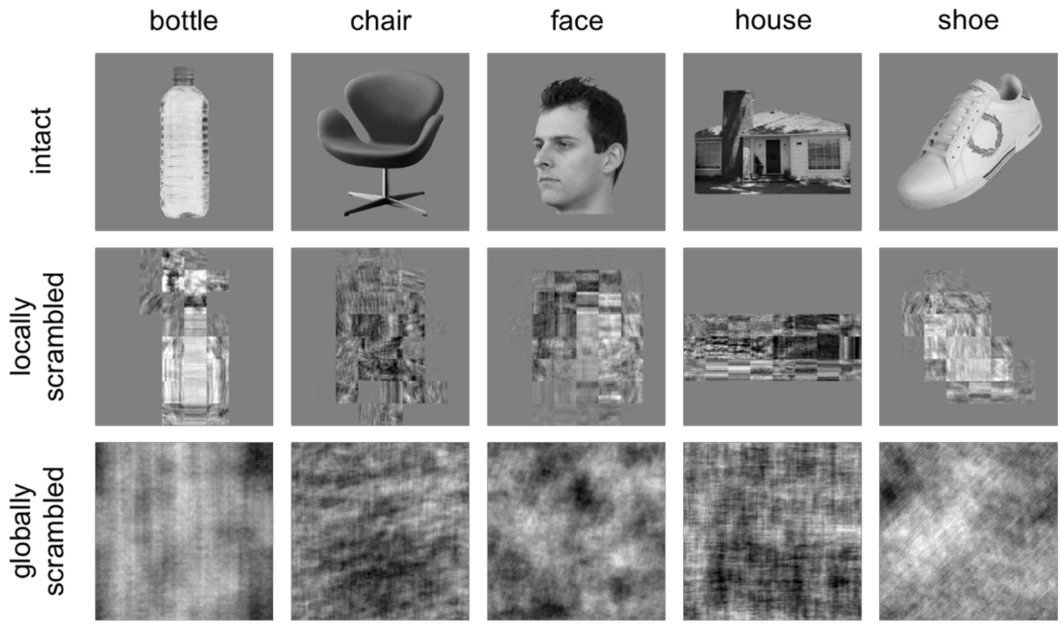

Figure 4.1 *Exemplars of intact, locally scrambled and globally scrambled images from the different object categories.*

4.3.2. 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.

4.3.3. 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. 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 box-car 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 (MNI-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 Figure 4.3 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 (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.

4.4. Results

First, we conducted a behavioural experiment to determine how image scrambling affected the categorical and semantic information that the images conveyed (Figure 2.2). 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.

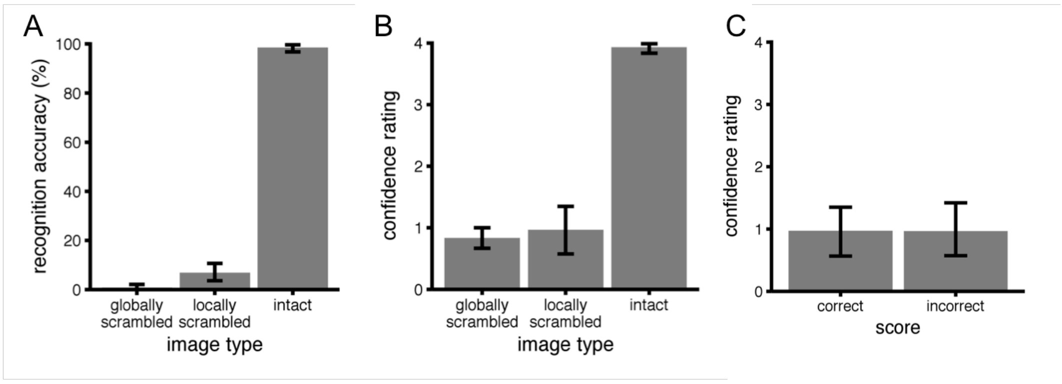

Figure 4.2 *Results of semantic naming task. (A) Mean accuracy scores for each image type. (B) Mean confidence ratings for correct responses for each image type. (C) Mean confidence ratings for correct and incorrect responses to locally scrambled images. For all panels, error bars reflect bootstrapped 95% confidence intervals.*

Next, we measured patterns of ventral response to intact, locally scrambled and globally scrambled images from different object categories. Figure 4.3 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.

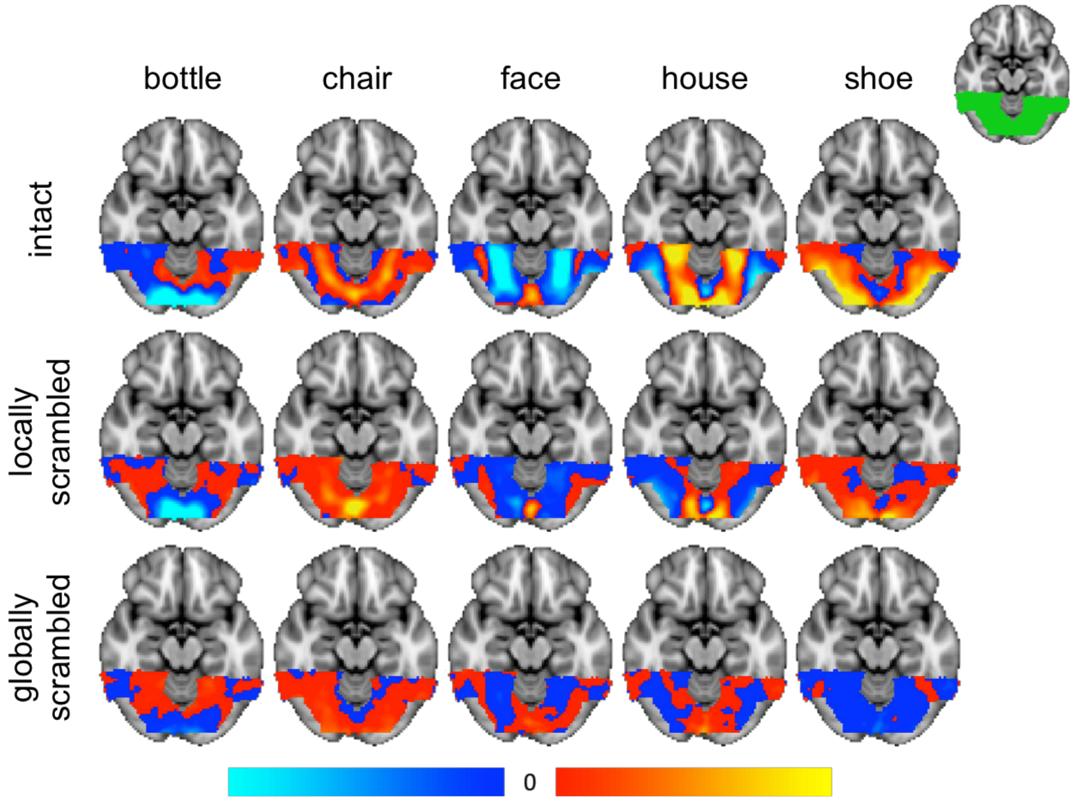

Figure 4.3 *Patterns of response to different object categories across the ventral visual pathway with intact, locally scrambled and globally scrambled conditions. Red/yellow and blue/light blue colours represent positive and negative fMRI responses, respectively, relative to the mean response across 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.*

To quantify the reliability of these patterns across participants, we used correlation-based MVPA. Figure 4.4A 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 Figure 4.4B) using an ANOVA with Comparison (within-category, between-category) and Category (bottle, chair, face, house, shoe) as the main factors.

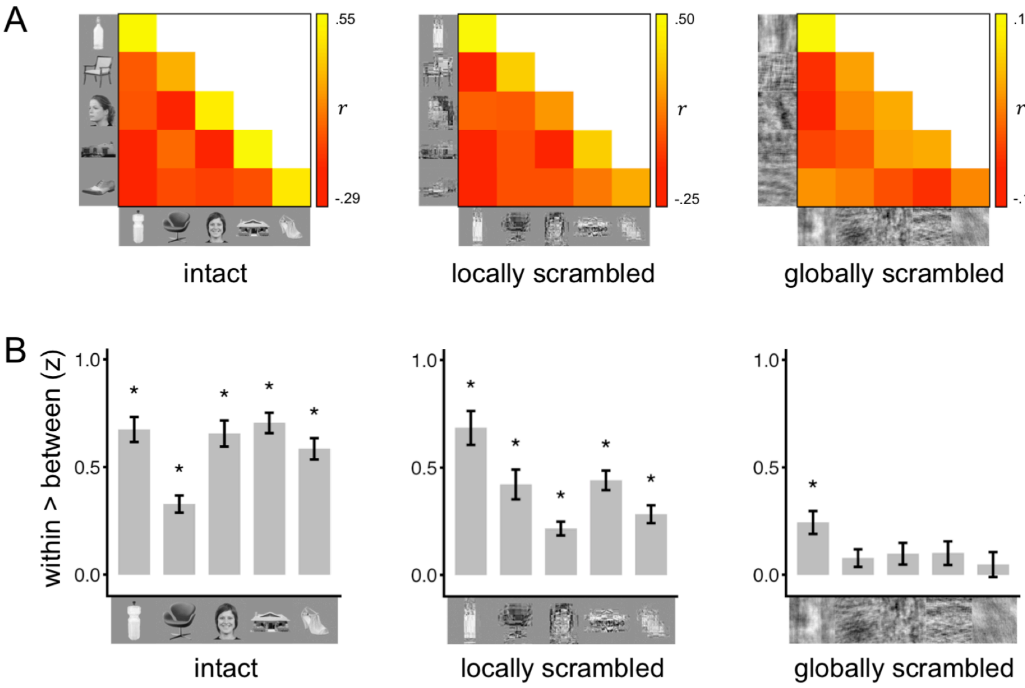

Figure 4.4 *Distinct patterns of response to different object categories with intact and scrambled images. (A) Similarity matrices showing within- and between-category correlations of neural patterns of response to intact, locally scrambled and globally scrambled object categories. (B) Bar graphs showing within-category minus between-category correlations. Error bars represent ±1 standard error of the mean. * denotes p < .05.*

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), chairs (t(21) = 8.25, p < .0001), faces (t(21) = 10.82, p < .0001), houses (t(21) = 14.85, p b .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).

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 Figure 4.4A). Scatter plots for each comparison are shown in Figure 4.5. 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.

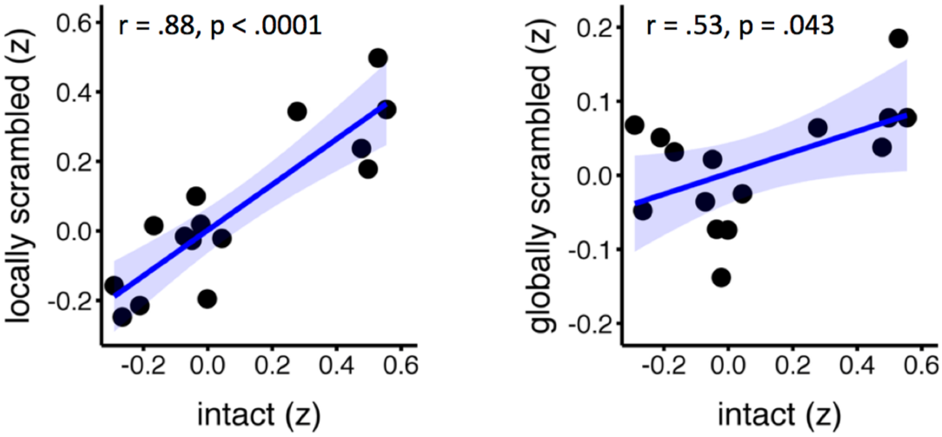

Figure 4.5 *Similar patterns of response to intact and scrambled images. Scatter plots show the correlation between the similarity matrices in Figure 4.4A. Significant positive correlations were found between matrices for intact and each of the scrambled image types, particularly the locally scrambled condition. Shaded regions indicate 95% confidence intervals for the best-fit regression lines.*

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 (Figure 4.6). This analysis suggests 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.

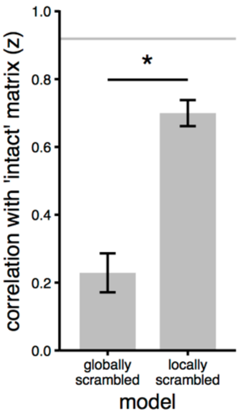

Figure 4.6 *Bar graph showing how variance in intact responses was accounted for by locally and globally scrambled responses. Grey line represents noise ceiling, which estimates the explainable variance in a dataset given the noise across participants. Responses to intact images were significantly better 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 (Figure 4.7). 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 (Table A.2.1), as indicated by a significant effect of Comparison (within-category > between-category correlations).

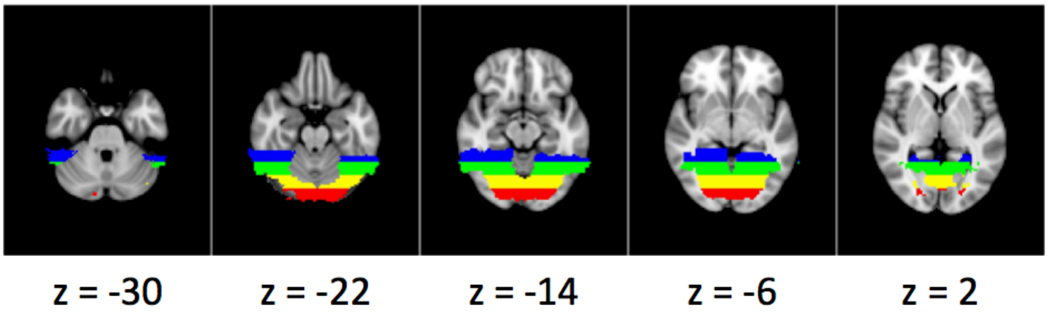

Figure 4.7 *Anatomical subdivisions of ventral stream mask along posterior-anterior axis. Each subdivision is identical in length along this axis.*

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 Table A.2.1).

Recent studies have shown retinotopic organization in more anterior regions of the ventral visual pathway (Arcaro et al., 2009; Wandell et al., 2007; 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, VO1/2, 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 Table A.2.1).

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 Figure 4.8). 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.

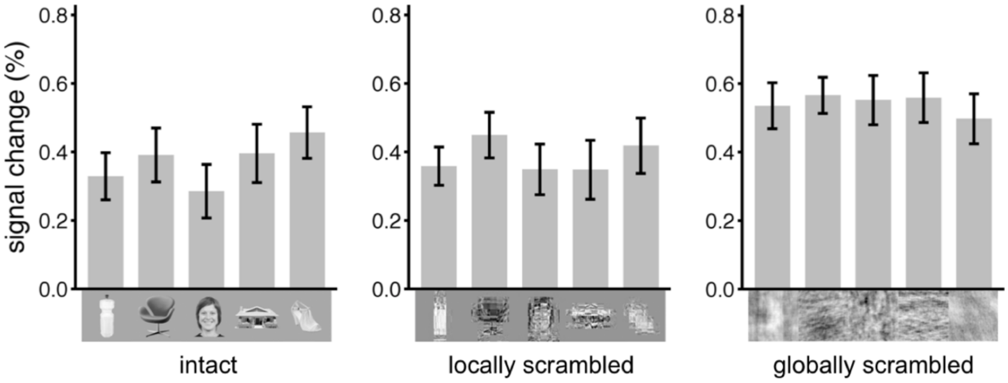

Figure 4.8 *Average signal change across the ventral stream to each category in each image type. Globally scrambled images evoked more activity than the locally scrambled and intact images. Error bars represent ±1 standard error of the mean.*

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. 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.

4.5. 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 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 (Andrews et al., 2015; Rice et al., 2014; Watson et al., 2014). 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 (Connolly et al., 2012; Kriegeskorte, Mur, Ruff, et al., 2008; Naselaris et al., 2009). 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 & 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 (Golomb & Kanwisher, 2012; Uri Hasson et al., 2002; Levy et al., 2001). Indeed, a number of studies have investigated the sensitivity of the ventral stream to shape information (Bracci, Ritchie, & Op de Beeck, 2017; Cant & Xu, 2012; Drucker & Aguirre, 2009; Haushofer, Livingstone, & Kanwisher, 2008; Kayaert, Biederman, & Vogels, 2003; Kourtzi & Kanwisher, 2001; Op de Beeck, Wagemans, & Vogels, 2001; Watson, Young, et al., 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 (Arcaro et al., 2009; Silson, Chan, Reynolds, Kravitz, & Baker, 2015; Wandell et al., 2007; 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 (Hodges, Patterson, Oxbury, & Funnell, 1992; Warrington, 1975). 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).

Chapter 5. Selectivity for the Mid-Level Properties of Faces and Places in the Fusiform Face Area and Parahippocampal Place Area

This chapter is adapted from: Coggan, D. D., Baker, D. H., & Andrews, T. J. (in press). Selectivity for the mid-level properties of faces and places in the fusiform face area and parahippocampal place area. *European Journal of Neuroscience.*

5.1. Abstract

Regions in the ventral visual pathway, such as the fusiform face area (FFA) and parahippocampal place area (PPA), are selective for images from specific object categories. Yet images from different object categories differ in their image properties. To investigate how these image properties are represented in the FFA and PPA, we compared neural responses to locally scrambled images (in which mid-level, spatial properties are preserved) and globally scrambled images (in which these mid-level, spatial properties are not preserved). There was a greater response in the FFA and PPA to images from the preferred category relative to their non-preferred category for the scrambled conditions. However, there was a greater selectivity for locally scrambled compared to globally scrambled images. Next, we compared the magnitude of fMR adaptation to intact and scrambled images. fMR-adaptation was evident to locally scrambled images from the preferred category. However, there was no adaptation to globally scrambled images from the preferred category. These results show that the

selectivity to faces and places in the FFA and PPA is dependent on mid-level properties of the image that are preserved by local scrambling.

5.2. Introduction

In previous chapters, we showed that category-selective responses across the ventral stream can be explained by more basic properties of the image. However, this may not be the case for sub-regions of the ventral stream that have shown the strongest category-selectivity, namely the FFA and PPA. Evidence that these regions are sensitive to low-level image properties is shown by higher responses to low-level properties (such as orientation and spatial frequency) that are typical of the preferred category (Rajimehr et al., 2011; Nasr and Tootell, 2012; Goffaux et al., 2016). Other studies have used Fourier scrambled images to investigate selectivity to low-level properties in these regions (Andrews et al., 2010; Rossion et al., 2012). The rationale for using scrambled images is that they contain many of the image properties found in intact images, but do not convey the same categorical or semantic information, thus providing a dissociation between higher-level and lower-level information. These studies have found mixed results. One study found selectivity to scrambled houses in PPA, but not to scrambled faces in the FFA (Andrews et al., 2010). However, another study found selectivity to scrambled faces in the FFA and other face-selective regions (Rossion et al., 2012).

The aim of this study was to use different methods of image scrambling to understand which image properties are important in the neural representations found in category-selective regions. To address this question, we compared the neural response in the FFA and PPA to intact images of faces and places with locally scrambled and globally scrambled versions of these images (Figure 5.1). Globally scrambled images were generated using a typical Fourier-scramble, i.e. keeping the global power of each two-dimensional frequency component constant while randomizing the phase of the components. Locally scrambled images were generated by windowing the original image into an 8 × 8 grid in image space and applying a phase-scramble to each window independently. A key difference between these methods of scrambling is that local scrambling preserves the low-level properties in their approximate original spatial location. This preserves some of the mid-level properties (such as the spatial envelope - the region of the image taken up by the object), which may play an important role in the representation of objects. In Chapter 4, we found that both global- and local-scrambling rendered the images unrecognizable (Coggan, Liu, et al., 2016). Despite the fact that images were unrecognizable, locally scrambled images, but not globally scrambled images, were found to elicit similar category-selective patterns of response across the ventral visual pathway (Coggan, Liu, et al., 2016). More recently, Long and colleagues (Long et al., 2018) showed that images that preserve mid-level properties of objects, but were not recognizable, elicited patterns of neural response to variation in animacy and real-world size that were comparable to intact objects.

Here, we ask whether there is a difference in the magnitude of response to locally scrambled and globally scrambled faces and places in the FFA and PPA. If selectivity is more evident to faces and places when they have been locally scrambled compared with global scrambling, then this shows that the magnitude of response in these regions can be explained in part by a sensitivity to the mid-level properties preserved by locally scrambled images. If there is no difference between locally scrambled and globally scrambled images, then this suggests that the selectivity found in scrambled images is due to the amplitude spectrum of the image. We also compared adaptation to faces and places with locally scrambled and globally scrambled images. The basis of fMRI adaptation is that repetition of a stimulus causes a reduction in the neural response, which leads to a lower fMRI signal, as explained in Section 2.3.2 (Grill-Spector and Malach, 2001; Avidan et al., 2002; Epstein et al., 2003; Andrews and Ewbank, 2004; Ewbank et al., 2005; Grill-Spector et al., 2006; Andrews et al., 2010, 2016; Psalta et al., 2014). Brain regions selective for a particular stimulus property will show greater adaptation (i.e. signal reduction) for a repeated stimulus than for a sequence in which the stimuli vary, whereas non-selective regions will show similar responses regardless of the sequence. The sensitivity of the neural representation can therefore be compared for different manipulations of the stimulus. If the underlying neural representation is insensitive to a particular type of manipulation in the stimulus (i.e. local- or global-scrambling), the adaptation of the fMRI signal will be similar to that produced by unchanged (in this case, intact) images.

5.3. Methods

5.3.1. Participants

Twenty participants were recruited for the fMRI experiment (12 female, mean age = 29.0 years, median = 23, min = 16, max = 66, SD = 12.7). Participants were constituted by graduate students and staff of the Department of Psychology at the University of York, as well as members of the public responding to a participant mailing list held by York Neuroimaging Centre (YNiC). The study was approved by the YNiC Ethics Committee and adhered to the original wording of the Declaration of Helsinki. All participants reported that they had normal or corrected-to-normal vision and gave their informed, written consent.

5.3.2. Stimuli

24 face and 24 house images were taken from a database of objects (Rice et al., 2014). Images were gray-scale, superimposed on a mid-gray background, and had a resolution of 720x720 pixels. Face images originated from the Radboud face database (Langner et al., 2010). 6 face and 6 house images were selected for adaptation scans, with the remaining images used in a localizer scan. For experimental scans, two different phase-scrambled versions of each image were generated. Global-scrambling 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. Local-scrambling involved windowing the original image into an 8x8 grid in image space and applying a phase-scramble to each window independently. Images subtended a maximum retinal angle of approximately 15° and were viewed on a screen at the rear of the scanner via a mirror placed immediately above the participant's head. Examples of the images are shown in Figure 5.1. The images used in this study have been validated by a behavioural study in Chapter 4 (Coggan, Liu, et al., 2016) in which participants were asked to name each image. The results of the naming task show that accuracy was at ceiling for intact images. However, local- and global-scrambling renders the images unrecognizable.

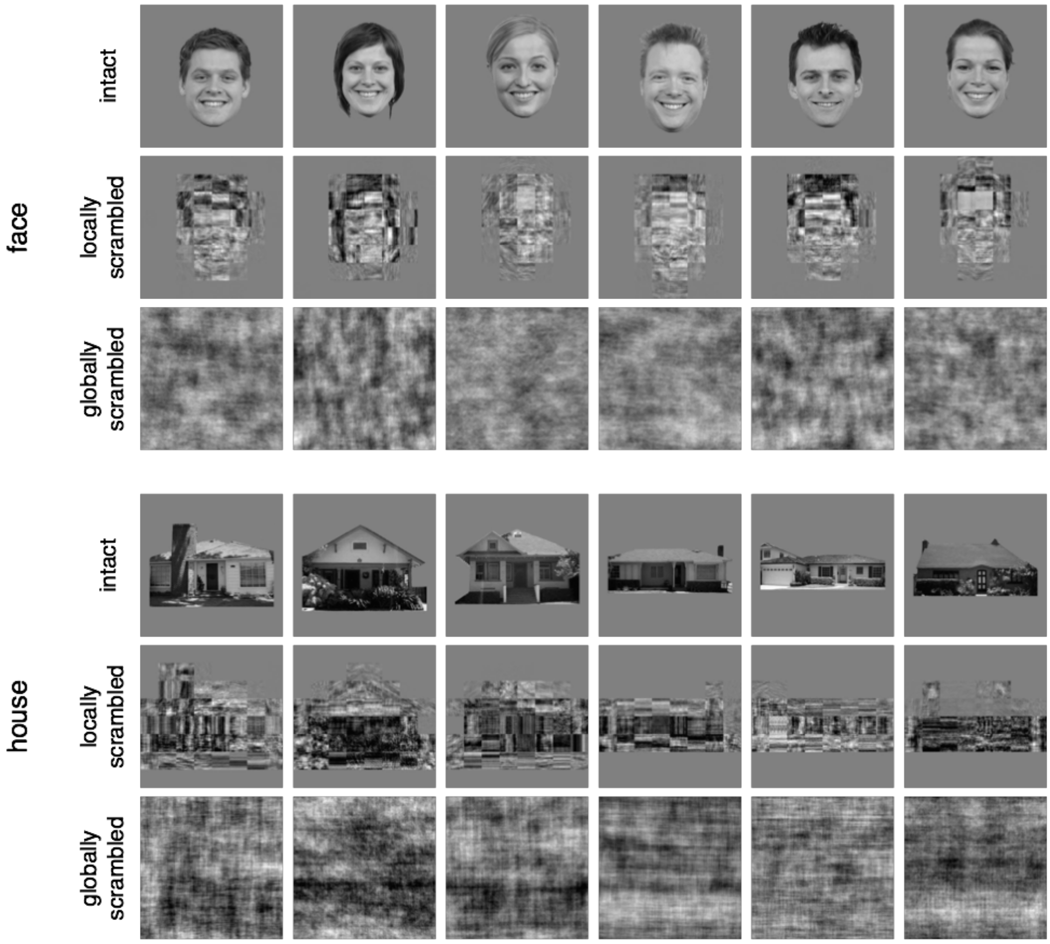

Figure 5.1 *Stimulus set containing intact, locally scrambled and globally scrambled versions of 6 face and 6 house images.*

5.3.3. Design and Procedure

There were 12 conditions in the experimental scan: 2 categories (face, house) x 3 image types (intact, locally scrambled, globally scrambled) x 2 adaptation sequences (same image, different images). The experiment was divided into 3 scan runs each lasting 8 minutes, with globally scrambled images presented in the first run, locally scrambled images presented in the second run and intact images presented in the third run. Scrambled images were presented before the intact images to prevent subjects becoming aware of the categories of the images during the presentation of the scrambled stimuli.

Images were presented in 6 s blocks. In each stimulus block, 6 images from a condition were presented for 800 ms with a 200 ms inter-stimulus-interval. This was followed by a fixation cross for 9 s. There were 8 repetitions of each condition in each scan.

To maintain attention, participants were instructed to press a button when a red dot appeared on any of the images. Subjects responded with a mean response latency of 423 ms (SEM = 10 ms). The number of correct responses was at ceiling for intact (mean = 99.0%, SEM = 0.4%), locally scrambled (mean = 99.5%, SEM = 0.2%) and globally scrambled (mean = 100%, SEM = 0%) conditions. Response latencies were entered into a one-way analysis of variance (ANOVA), which showed no effect of level of scrambling ($F(2,34) = 0.62$, $\eta^2 = .03$, $p = .5460$).

5.3.4. Data Acquisition and Analysis

fMRI data were acquired with a GE 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. 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 from the localizer and experimental scans were initially analyzed 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.

A localizer scan was performed after the experimental scan to localize the FFA and PPA in each individual. This involved a block-design paradigm with the same temporal parameters as the adaptation scans. Intact faces and houses were presented in alternate blocks, with 6 repetitions of each category. Images were different to those used in the adaptation experiment. Face- and place-selective voxels were identified using face>house and house>face contrasts, respectively. The resulting statistical maps were thresholded at z > 2.3. Within the anatomical location of the FFA and PPA, a flood-filling algorithm was used to define 100 spatially contiguous voxels in each hemisphere. The voxel with the highest z-score for each contrast was located. Then, voxels contiguous to that voxel with the highest z-score were iteratively added to generate a progressively larger mask. This process continued until 100 voxels had been reached, or there were no more significant contiguous voxels. It was not possible to identify the FFA and PPA in 3 participants (3 males aged 66, 33 and 24), so they were removed from further analyses. The final sample consisted of 17 subjects (12 female, mean age = 26.9 years, median = 23, min = 16, max = 56, SD = 9.9). The average location of the FFA and PPA across participants is shown in Table 5.1.

		peak coordinates				
ROI	hemisphere	x	y	z	voxels	Z
FFA	left	-42 (4.3)	-58 (1.0)	-23 (5.4)	88 (22)	4.0 (1.0)
	right	41 (4.2)	-53 (7.3)	-25 (4.8)	87 (15)	4.4 (1.2)
PPA	left	-27 (3.5)	-54 (5.7)	-14 (2.7)	95 (21)	4.8 (1.2)
	right	28 (3.2)	-55 (9.0)	-14 (5.1)	99 (5)	5.2 (1.3)

Table 5.1 *Group means (N=17) and standard deviations (in parentheses) for peak coordinates, number of voxels and average Z score for each ROI in the localization contrast. ROIs were transformed from individual-space in which the analysis was performed into MNI-ICBM152 2mm space for the purposes of this table.*

5.3.5. Experimental Scan

To compare the magnitude of response to each condition, parameter estimates were generated by regressing the hemodynamic response of each voxel against a boxcar function convolved with a single-gamma HRF. Responses from each voxel were averaged within each region of interest (ROI) and converted to percent signal change. A repeated-measures analysis of variance (ANOVA) was then used to determine the effect of ROI (FFA, PPA), Image Type (intact, locally scrambled, globally scrambled), Adaptation (same, different) and Preferred Category (FFA: preferred=face, non-preferred=house; PPA: preferred=house, non-preferred=face). An FDR correction for multiple comparisons (Benjamini & Hochberg, 1995) was applied to all post-hoc, pairwise comparisons. All comparisons were two-tailed. FSL's featquery was used to obtain signal change estimates in each ROI. From there, the ANOVA, post-hoc tests and plotting were all performed using R (https://www.r-project.org). The R code and signal change estimates are available at https://github.com/ddcoggan/p004. Statistical analyses were performed on the mean values across participants.

5.4. Results

First, the selectivity for the preferred object category (faces for FFA, houses for PPA) was measured with intact, locally-scrambled and globally-scrambled images. The magnitude of response to intact and scrambled faces and houses to the preferred and non-preferred categories is shown in Figure 5.2. There were main effects of Preferred Category (F(1,16) = 384.41, partial η^2 = .96, p < .0001) and Image Type (F(2,32) = 42.48, partial η^2 = .73, p < .0001). The effect of Preferred Category was due to a higher response to the preferred compared to the non-preferred stimulus with intact (t(16) = 18.94, mean difference = 0.78 [95% CI = 0.69, 0.86], Cohen's *d* = 4.59, p < .0001), locally-scrambled (t(16) = 7.55, mean difference = 0.19 [0.14, 0.25], Cohen's *d* = 1.83, p < .0001) and globally-scrambled (t(16) = 2.72, mean difference = 0.03 [0.01, 0.06], Cohen's *d* = 0.66, p = .0229) images.

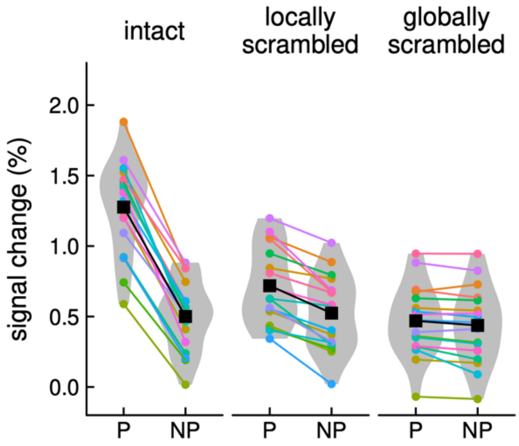

Figure 5.2 *Percent signal change in category-selective regions (collapsed across FFA and PPA) to intact, locally scrambled and globally scrambled images of their preferred (P) and non-preferred (NP) categories. Colours reflect different subjects, with the group mean shown in black. Grey surrounds are symmetrical kernel density estimates reflecting distribution of values.*

There was a two-way interaction between Preferred Category and Image Type (F(2,32) = 189.74, partial η^2 = .92, p < .0001). The interaction suggests that selectivity for the preferred category varies for different image types. To test this, we compared the difference between the preferred and non-preferred category for each image type. The difference between the preferred and non-preferred category was greater for intact images compared to both locally-scrambled (t(16) = 12.27, mean difference = 0.58 [0.48, 0.68], Cohen's *d* = 2.97, p < .0001) and globally-scrambled (t(16) = 18.14, mean difference = 0.74 [0.66, 0.83], Cohen's *d* = 4.40, p < .0001) images. However, there was also a bigger difference between the preferred and non-preferred stimulus for locally-scrambled compared to globally-scrambled images (t(16) = 5.38, mean difference = 0.16 [0.10, 0.23], Cohen's *d* = 1.30, p < .0001). There was no significant interaction between Preferred

Category, Image Type and ROI (F(2,32) = 1.38, partial η^2 = .08, p = 0.27). Consistent with the ROI analysis, Figure 5.3 shows the relative response to faces and houses is similar for intact and locally-scrambled images, but a different pattern of response to the globally-scrambled images (see Chapter 4, Figure 4.2).

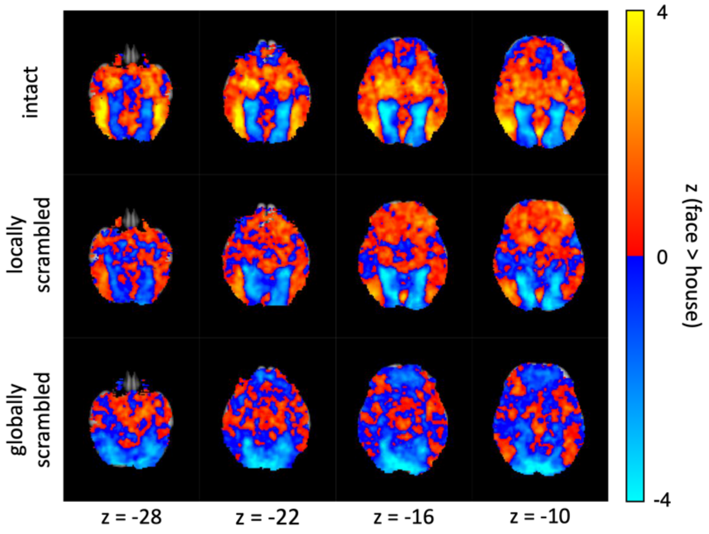

Figure 5.3 *Axial slices showing group-level z statistics for a face > house contrast for each image type in 2mm MNI space. For each image type, data were collapsed across 'different' and 'same' sequence types to form one parameter estimate per category. Red/yellow regions were more responsive to faces; blue regions were more responsive to houses.*

Next, we asked whether the FFA and PPA would show fMR-adaptation to intact, locally scrambled and globally scrambled images of preferred and non-preferred categories (Figures 5.4). There was a main effect of Adaptation (F(1,16) = 36.98, partial η^2 = .70, p < .0001) and a three-way interaction between Preferred Category, Image Type and Adaptation (F(2,32) = 11.49, partial η^2 = .42, p = .0002). This indicates that the level of adaptation varied with the preferred stimulus and level of scrambling (Figure 5.5).

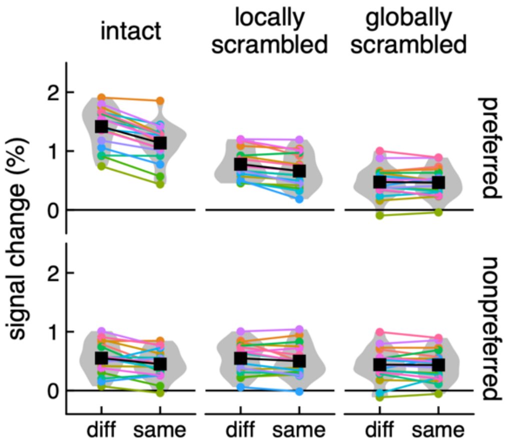

Figure 5.4 *Percent signal change in category selective regions (collapsed across FFA and PPA) in response to intact, locally scrambled and globally scrambled images of their preferred and non-preferred categories, presented as a sequence of different images or the same image repeated. Colours reflect different subjects, with the group mean shown in black.*

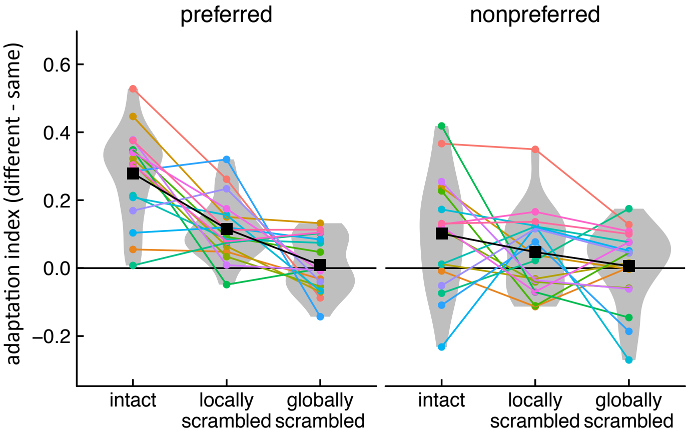

Figure 5.5 *Adaptation index (different – same) in category selective regions (collapsed across FFA and PPA) in response to intact, locally scrambled and globally scrambled images of their preferred and non-preferred categories. Colours reflect different subjects, with the group mean shown in black.*

Pairwise comparisons revealed significant adaptation (different > same) to the preferred category for intact (t(16) = 8.38, mean difference = 0.28 [0.21, 0.35], Cohen's *d* = 2.03, p < .0001) and locally scrambled images (t(16) = 5.09, mean difference = 0.12 [0.07, 0.16], Cohen's *d* = 1.24, p = .0002), but not to globally scrambled images (t(16) = 0.44, mean difference = 0.01 [-0.03, 0.06], Cohen's *d* = 0.11, p = .7199). The magnitude of the adaptation to the preferred category was bigger for intact images compared to locally-scrambled (t(16) = 4.39, mean difference = 0.16 [0.08, 0.24], Cohen's *d* = 1.06, p = .0015) and globally-scrambled (t(16) = 6.63, mean difference = 0.27 [0.18, 0.36], Cohen's *d* = 1.61, p < .0001) images. The magnitude of the adaptation to the preferred category was bigger for locally-scrambled images compared to globally-scrambled images (t(16) = 3.02, mean difference = 0.11 [0.03, 0.18], Cohen's *d* = 0.73, p = .0182). In contrast, there was only an effect of adaptation for the non-preferred category with intact images (t(16) = 2.43, mean difference = 0.10 [0.01, 0.19], Cohen's *d* = 0.59, p = .0465) and no significant effect for locally-scrambled (t(16) = 1.60, mean difference = 0.05 [-0.02, 0.11], Cohen's *d* = 0.39, p = .1673) or globally-scrambled (t(16) = 0.20, mean difference = 0.01 [-0.05, 0.07], Cohen's *d* = 0.02, p = .8622) images. Finally, there was no significant interaction between Preferred Category, Image Type, Adaptation and ROI (F(2,32) = 0.90, partial η^2 = .05, p = .42), again demonstrating that these effects generalise across regions.

Finally, we investigated whether the results described above were inherited from responses to the images in early visual cortex (Figure 5.6). To address this, we registered individual-level data into a standard space (MNI152) and restricted our analysis to a V1 mask taken from a probabilistic atlas of retinotopic regions (Wang et al., 2015). A three-way repeated-measures ANOVA revealed main effects of Image Type (F(2,32) = 27.6, partial η^2 = .63, p < .001), Category (F(1,16) = 28.7, partial η^2 = .64, p < .001) and Adaptation (F(1,16) = 25.2, partial η^2 = .61, p < .001). In contrast to the responses in higher-level regions, the effect of Image Type in V1 was due to higher responses to globally-scrambled compared to both locally-scrambled (t(16) = 2.94, mean difference = 0.27 [0.07, 0.46], Cohen's *d* = 0.71, p = .0096) and intact (t(16) = 6.80, mean difference = 0.71 [0.49, 0.94], Cohen's *d* = 1.65, p < .0001) images. There was a higher response to locally-scrambled compared to intact images (t(16) = 4.71, mean difference = 0.45 [0.25, 0.65], Cohen's *d* = 1.14, p = .0004). The effect of category was due a higher response to houses compared to faces (t(16) = 5.36, mean difference = 0.17 [0.10 – 0.24], Cohen's *d* = 1.30, p < .0001). The effect of Adaptation was due to sequences of different images eliciting greater response than sequences of the same image. There were no significant interactions between Image Type and Adaptation (F(2,32) = 2.09, partial η^2 = .12, p = .139), Category and Adaptation (F(1,16) = 3.29, partial η^2 = .17, p = .088) or between Image Type, Category and Adaptation (F(2,32) = 0.13, partial η^2 = .01, p = .878). Taken together, this shows that adaptation in V1 was not significantly different across image type or category.

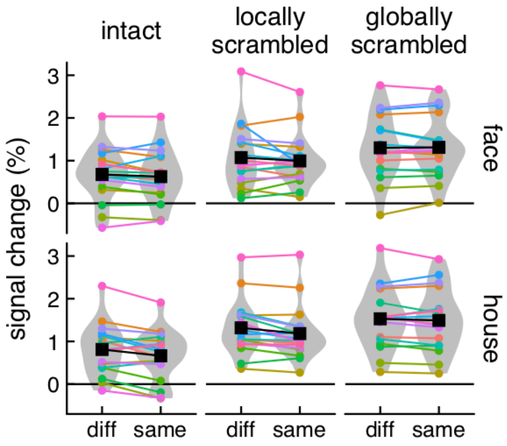

Figure 5.6 *Signal change in V1 in response to sequences of different and same images for each image type and category. Adaptation (different – same) changed very little across categories or image types. Colours reflect different subjects, with the group mean shown in black.*

5.5. Discussion

The aim of this study was to explore the sensitivity of category-selective regions in the ventral stream to low-level image properties. To test this, neural responses to intact, locally scrambled and globally scrambled images of faces and houses were compared in the face-selective FFA and place-selective PPA. The rationale for using scrambled images is that they contain many of the image properties found in intact images, but do not convey the same categorical or semantic information, thus providing a dissociation between higher-level and lower-level information. However, in this study we also compared the response to locally scrambled and globally scrambled images. The major difference between these image types is that spatial properties of the image are preserved in the locally scrambled images, but not in the globally scrambled images. Previous studies have found selectivity to globally scrambled images in the FFA and PPA (Andrews et al., 2010; Rossion et al., 2012). The key finding from this study is that selectivity and adaptation to the preferred category in the FFA and PPA were greater for the locally scrambled images compared to the globally scrambled images.

These results imply that the selectivity to faces and places in the FFA and PPA is to some extent determined by the spatial properties of the image. These findings complement those of Chapter 4 in which the pattern of neural response to intact images is shown to be more similar to locally scrambled compared to globally scrambled images. These findings also fit with previous studies that have demonstrated selectivity in higher-level regions of the ventral stream to spatial properties of the image (Bracci & Op de Beeck, 2016; Cichy et al., 2013; Golomb & Kanwisher, 2012; Levy et al., 2001; Ponce, Sturmfels, & Trager, 2017; Watson, Young, et al., 2016). More generally, these results are consistent with previous studies that have shown patterns of response in high-level visual regions are sensitive to the image properties (Rice et al., 2014; Watson, Young, et al., 2016; Xu, Yue, Lescroart, Biederman, & Kim, 2009; Yue, Tjan, & Biederman, 2006). For example, patterns of response in the fusiform gyrus to faces can be predicted by their image properties (Rice et al., 2014). Moreover, equivalent changes in the image statistics that result in either a change in identity or no change in identity lead to an equivalent release from adaptation in regions such as the occipital face area (OFA) and FFA (Xu et al., 2009; Yue et al., 2006).

The selectivity and adaptation for the preferred category was greater in intact images compared to locally scrambled images. One possible explanation for this difference is that the neural representation is selective for higher-level, semantic information about the image that is only available from the intact images (Kanwisher, 2010). However, an alternative possibility is that unexplained variance might reflect image properties disrupted by the scrambling process. An important feature of intact images is the strong statistical dependencies between features, such as specific combinations of spatial frequency and orientation at particular locations in the image. Indeed, the behavioural sensitivity to these regularities in intact objects suggests that they play an important role in differentiating between different classes of images (Loschky et al., 2007; Loschky & Larson, 2010). It is possible that these properties also contribute to the patterns of response in category-selective regions. When evaluating these possibilities, it is important to recognize that high-level and low-level contributions to the observed representational structure are not mutually exclusive. The extraction of any high-level features depends on the availability of relevant low-level features being preserved in the scrambled stimuli.

Although adaptation was most evident to images from the preferred category, we also found significant adaptation to intact images from the non-preferred category. This finding is relevant to recent accounts that have attempted to explain the organization of the occipital–temporal cortex (Behrmann & Plaut, 2013). The domain-specific approach suggests that discrete cortical regions are selective for the processing of specific categories of objects (Kanwisher, 2010). In contrast, the domain-general approach, suggests a distributed and overlapping representation of visual information along the occipital–temporal lobe (Haxby et al., 2001). Neuropsychological studies are often used as evidence for a domain-specific representation (McNeil & Warrington, 1993; Moscovitch et al., 1997). Our finding of adaptation to the non-preferred object category is consistent with previous studies that have found adaptation to non-preferred stimuli in category-selective regions (Ewbank et al., 2005). This suggests that the representation of objects and places is not restricted to those regions that respond maximally but is distributed across the ventral visual pathway. However, it is also important to reiterate that the magnitude of the adaptation was much greater for the preferred compared to the non-preferred category.

Finally, we asked whether the pattern of results found in higher-level regions reflected responses at early stages of processing. The pattern of response in V1 was quite different to that found in the category-selective regions. We found the highest response to globally scrambled images, which presumably reflects differences in the amount of the visual field that was stimulated (Grill-Spector, Kushnir, Edelman, Itzchak, & Malach, 1998). Although there was adaptation to repetitions of the same object, this was not significantly different for intact or scrambled images. Together, these results demonstrate the responses in the FFA and PPA are emergent properties of the visual system.

In conclusion, we have shown that the selectivity to objects in category-selective regions is also evident to locally scrambled objects in which the spatial properties of the image are preserved, but is less evident to globally scrambled objects in which spatial properties are disrupted. This suggests that the neural representation in high-level visual cortex is particularly sensitive to the spatial properties of the stimulus. Nevertheless, it is clear that the selectivity and adaptation demonstrated by scrambled images does not explain all of the variance in the intact images. Further studies will be needed to understand the relative role of image properties not preserved by scrambling and higher-level semantic properties in the neural representation of category-selective regions.

Chapter 6. The Emergence of Object-Selectivity in Early Visual Areas (V1 – V3)

This chapter is adapted from: Coggan, D. D., Allen, L. A., Farrar, O. R. H., Gouws, A. D., Morland, A. B., Baker, D. H., & Andrews, T. J. (2017). The emergence of object-selectivity in early visual areas (V1-V3). *Scientific Reports, 7,* 2444.

6.1. Abstract

High-level regions of the ventral visual pathway respond more to intact objects compared to scrambled objects. The aim of this study was to determine if this selectivity for objects emerges at an earlier stage of processing. Visual areas (V1–V3) were defined for each participant using retinotopic mapping. Participants then viewed intact and scrambled images from different object categories (bottle, chair, face, house, shoe) while neural responses were measured using fMRI. Our rationale for using scrambled images is that they contain the same low-level properties as the intact objects but lack the higher-order combinations of features that are characteristic of natural images. Neural responses were higher for scrambled than intact images in all regions. However, the difference between intact and scrambled images was smaller in V3 compared to V1 and V2. Next, we measured the spatial patterns of response to intact and scrambled images from different object categories. We found higher within-category compared to between category correlations for both intact and scrambled images demonstrating distinct patterns of response. Spatial patterns of response were more distinct for intact compared to scrambled images in V3, but not in V1 or V2. These findings demonstrate the emergence of selectivity to natural images in V3.

6.2. Introduction

A characteristic of high-level visual regions is their selectivity for intact images (Andrews et al., 2010; Grill-Spector et al., 1998; Malach et al., 1995). At early stages of processing (V1), there are greater responses to scrambled compared to intact images (Grill-Spector et al., 1998). In contrast, the responses in high-level visual cortex are greater for intact compared to scrambled images. The selectivity for intact images is also evident in the spatial pattern of response of high-level regions of the ventral pathway. In Chapter 4, more distinct patterns of neural response (defined by higher within- compared to between-category correlations) were found to intact compared to scrambled images. An important feature of intact images is the strong statistical dependencies between features, such as location-specific combinations of orientation and spatial frequency. Indeed, the behavioural sensitivity to the regularities that occur in intact objects suggests that these properties are critical for differentiating between different classes of images (Geisler, 2008; Oppenheim & Lim, 1981; Sigman, Cecchi, Gilbert, & Magnasco, 2001; Thomson, 1999; Vogels, 1999).

The aim of this study is to determine at what stage these statistical properties of intact objects emerge in the ventral visual pathway. Recent studies have shown that neurons in V2, but not in V1, respond selectively to synthetic textures that are based on the higher-order statistical properties found in natural images (Freeman & Simoncelli, 2011; Freeman, Ziemba, Heeger, Simoncelli, & Movshon, 2013). Patterns of response in V2 are better able discriminate these naturalistic textures than control textures that are not based on natural images (Ziemba, Freeman, Movshon, & Simoncelli, 2016). In the current study, we compare the response to images of objects and to scrambled versions of objects in early visual areas (V1-V3). Our aim was to determine whether these early visual areas showed selectivity to the statistical properties found in natural images. Our hypothesis was that, if neurons in early visual areas are selective for statistical regularities found in intact objects, there should be either a greater response or a more distinct pattern of response to intact compared to scrambled images.

6.3. Methods

6.3.1. Stimuli

180 images of three object categories (face, bottle, house) were taken from an object image stimulus set (Rice et al., 2014). All images were gray-scale, superimposed on a mid-gray background, and had a resolution of 680x680 pixels. A scrambled version of each image was created by applying a Fourier phase-scramble to different spatial regions of the image, described as local scrambling in Chapters 4 and 5. This involved windowing each image into an 8x8 grid and phase-scrambling the contents of each window independently. This process preserves the spatial extent of the images. Images subtended a maximum retinal angle of approximately 15° and were viewed on a screen at the rear of the scanner via a mirror placed immediately above the participant's head. Examples of the stimuli are shown in Figure 6.1.

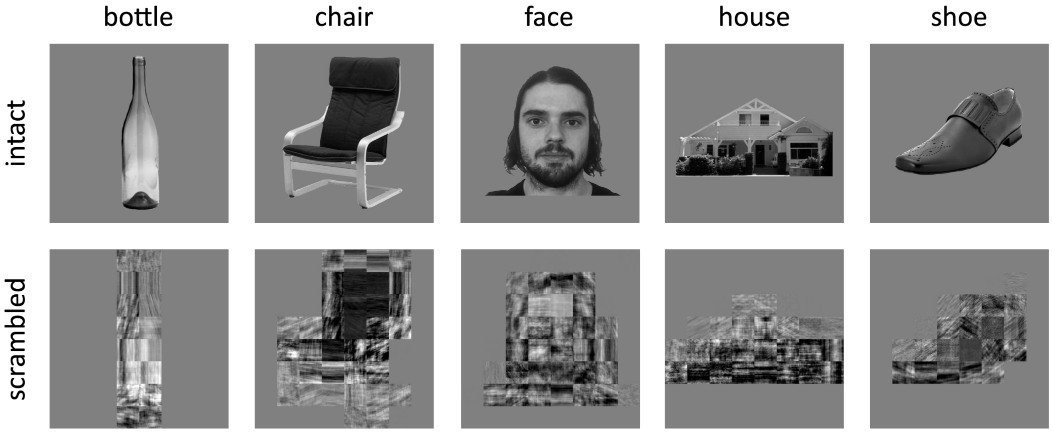

Figure 6.1 *Exemplars of intact and scrambled images from the different object categories.*

6.3.2. Participants

Twenty-one participants took part in the fMRI experiment (10 male, mean age = 26.3, SD = 6.0 years). All participants had normal or corrected-to-normal vision. Each gave their informed, written consent and the study was approved by the York Neuroimaging Centre (YNiC) Ethics Committee and adhered to the Declaration of Helsinki.

6.3.3. Design and Procedure

There were 10 conditions: 5 categories (bottle, chair, face, house, shoe) x 2 image types (intact, scrambled). Images were presented in 6 s blocks. In each block, 6 images from a condition were presented individually for 800 ms, with a 200 ms inter-stimulus-interval. This was followed by a fixation cross lasting 9 s. There were 6 repetitions of each condition in the scan. To maintain attention participants were instructed to press a button on a response box whenever a red dot appeared on an image, which occurred once in each block. On average, subjects responded to 99.3% (SEM = 0.04%) of red dot images, with a mean reaction time of 420 ms (SEM = 14 ms). There was no significant difference in the number of hits between intact (mean = 99.4%, SEM = 0.3%) and scrambled (mean = 99.2%, SEM = 0.5%) conditions (t(20) = 0.37, ns). There was also no significant difference in the response latencies between intact (mean = 417 ms, SEM = 15 ms) and scrambled (mean = 425 ms, SEM = 14 ms) conditions (t(20) = 1.56, ns).

6.3.4. Data Acquisition

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. 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 analyzed 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.

6.3.5. Region of Interest Localization

Visual areas were defined in a separate scan session (TR, 3000 ms; TE, 30 ms; voxel size, 2x2x2 mm^3; flip angle, 90°; matrix size, 96x96x 39; FOV, 19.2 cm) with a 16-channel head coil to improve signal-to-noise in the occipital lobe using either ring and wedge type stimuli or population receptive field techniques, as described in Section 2.5.3. Wedges rotated counter-clockwise about a red fixation cross. Ring stimuli expanded about fixation. Both wedges and rings were high contrast checkerboard stimuli that flickered at a rate of 6 Hz. Each scan contained eight cycles of wedges/rings, with 36 s per cycle, traversing a circular region of radius 14.3°. Participants maintained fixation throughout the scan. Visual area boundaries between V1/V2 and V2/V3 (dorsal and ventral) were defined by the phase reversals in the polar angle representations on inflated representations of the visual cortex (Figure 6.2). Visual field eccentricity representations were used to restrict the ROI to the location of the stimulus, i.e. the central 15° of visual angle. Functional data from the main experimental scan were aligned to a high-resolution T1-anatomical image that was segmented into gray matter and white matter.

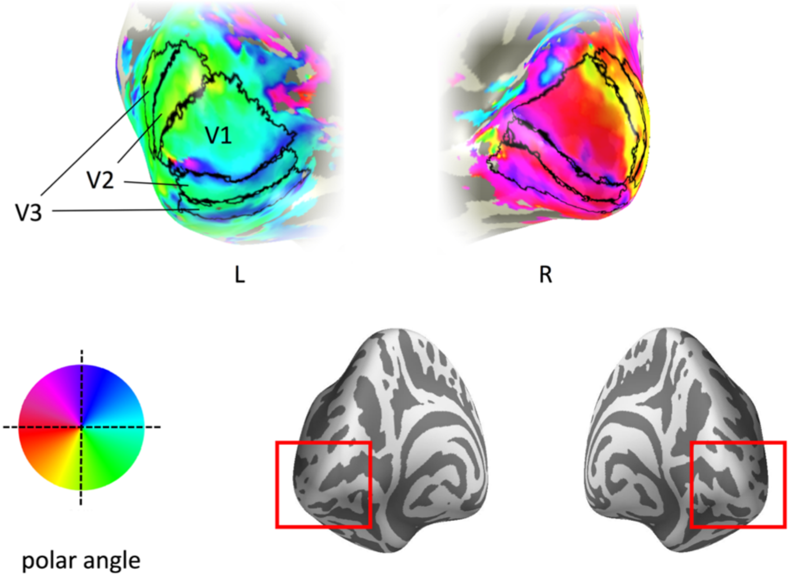

Figure 6.2 *Early visual cortical regions for a representative subject. Visual areas are superimposed onto the occipital lobe – see red insert on the posterior view of the inflated brain. Colour maps indicate the preferred polar angle.*

6.3.6. Data Analysis

To compare the magnitude of response to different stimulus conditions, 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. The responses from each voxel were then averaged within each ROI and converted from units of image intensity to % signal change. A repeated measures ANOVA was then used to determine the effect of ROI (V1, V2, V3) and Image Type (intact, scrambled).

To compare the spatial patterns of neural response, parameter estimates were generated for odd and even runs of each condition by regressing the hemodynamic response of each voxel against a box-car function convolved with a single-gamma HRF. Parameter estimates were normalized by subtracting the mean response per voxel across all conditions (odd and even, intact and scrambled). These data were then submitted to a within-subjects, correlation-based multivariate pattern analysis (Haxby et al., 2001,

2014) implemented using the PyMVPA toolbox (http://www.pymvpa.org/, Hanke et al., 2009). This allowed us to compare spatial patterns of response to all combinations of objects. For within-category comparisons, the correlation between responses in odd and even runs was used. For between-category comparisons, the mean correlation across odd-even and even-odd contrasts was used. A Fisher's Z-transformation was then applied to the correlations prior to further statistical analyses. If there are distinct patterns for each object category, there should be a higher correlation in the spatial pattern of response for within-category compared to between-category comparisons.

6.4. Results

First, we asked whether the overall neural response in V1, V2 and V3 could distinguish intact and scrambled images. To address this question, we measured the % signal change in each region to intact and scrambled images (Figure 6.3). We then performed a 2-way ANOVA with Region (V1, V2, V3) and Image Type (intact, scrambled) as factors. There was a significant main effect of Image Type (F(1,20) = 54.67, p < .0001) and a significant interaction between Region and Image Type (F(2,40) = 10.83, p = .0002). Pairwise comparisons revealed that scrambled images evoked more activity than intact images in V1 (t(20) = 6.46, p < .0001), V2 (t(20) = 7.55, p < .0001) and V3 (t(20) = 5.28, p = .0001). However, this difference was significantly smaller in V3 compared to both V1 (t(20) = 3.13, p = .0079) and V2 (t(20) = 5.54, p < .0001) (see Figure 6.3B). This is the cause of the interaction detected by the ANOVA analysis as there was no difference between V1 and V2 (t(20) = 0.97, ns).

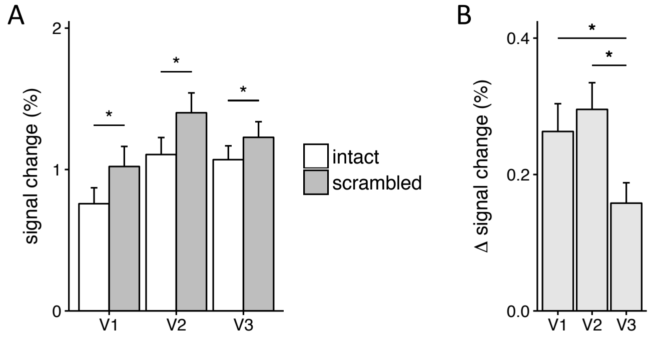

Figure 6.3 *Univariate results. (A) Magnitude of response to intact and scrambled images. Scrambled images evoked more activity than intact images in each visual area. (B) Differences in response to intact and scrambled images for each visual region. V3 showed a smaller difference in response to intact and scrambled images compared to V1 and V2. Error bars show ±1 SEM.*

** p < .05, FDR corrected.*

Next, we asked whether there were differences in the spatial patterns of response in V1, V2 and V3 to intact and scrambled images. To address this question, we first tested whether different intact and scrambled object categories evoked distinct and reliable patterns of fMRI response in regions V1, V2 and V3 (Figure 6.4). We compared the similarity of patterns of response to images from the same category (e.g. bottle vs. bottle) with the similarity of patterns to images of different categories (e.g. bottle vs. chair). Distinct, category-specific patterns of response are indicated by the within-category correlations being significantly greater than the between-category correlations (see Figure 6.4B).

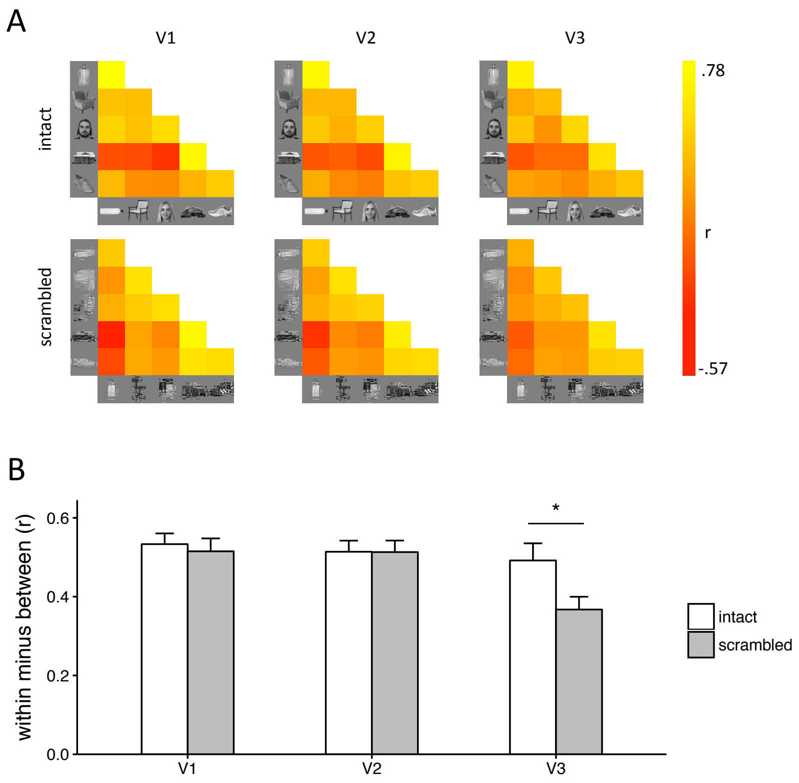

Figure 6.4 *Multivariate results. (A) Similarity matrices showing the correlation in patterns of neural response to all within-category and between-category comparisons. Within-category comparisons (e.g. bottle-bottle) are shown on the diagonal. (B) Bar graph showing the mean within-category and between-category correlations for intact and scrambled images across participants. There was a significant interaction between Comparison, Image Type and Region, which was due to more distinct (within>between) patterns of neural response to intact relative to scrambled images in V3. Error bars show ±1 SEM.*

** p < .05, FDR corrected.*

A 3-way repeated-measures ANOVA was performed with Comparison (within-category, between-category), Region (V1, V2, V3) and Image Type (intact, scrambled) as factors. There were main effects of Comparison (F(1,20) = 383.15, p < .0001) and Region (F(2,40) = 10.40, p = .0002). Although there was no effect of Image Type (F(1,20) = 1.24, p = .28), there was a significant three-way interaction between Comparison, Region and Image Type (F(2,40) = 5.28, p = .0093). This indicated that the distinctiveness of the category-specific patterns of response reflected by the effect of Comparison (within-category - between-category) differed across intact and scrambled images, depending on the visual region. Pairwise comparisons revealed that intact images evoked more distinct category-specific patterns than scrambled images in V3 (t(20) = 2.99, p = .020). This difference in the spatial pattern of response was not seen in V1 (t(20) = 1.17, ns) or V2 (t(20) = 0.71, ns). This shows that the spatial pattern of response to different object categories is more distinct for intact compared to scrambled images in V3.

Finally, we investigated the patterns of response similarity to different object categories across image type and region. Figure 6.5A shows all pairwise correlations across the different similarity matrices shown in Figure 6.4. This appears to show higher correlations to the same image type (intact or scrambled). For example, in V1 the correlation with V2 and V3 for intact images was 0.99 and 0.96, respectively. In contrast, the correlation between intact and scrambled images in V1 was 0.63. To assess whether the higher correlation for the same image type was statistically significant, we generated two models: Image Type and Region (Figure 6.5B). These models were regressed onto each subject's matrix, generating a distribution of beta-weights for each model (Figure 6.5C). These weights were significantly above zero for Image Type (t(20) = 8.45, p < .0001), but not for Region (t(20) = 0.73, ns). Weights for Image Type were also significantly higher than those for Region (t(20) = 8.26, p < .0001). This shows that representational distances between the object categories across all visual areas was different for intact and scrambled images.

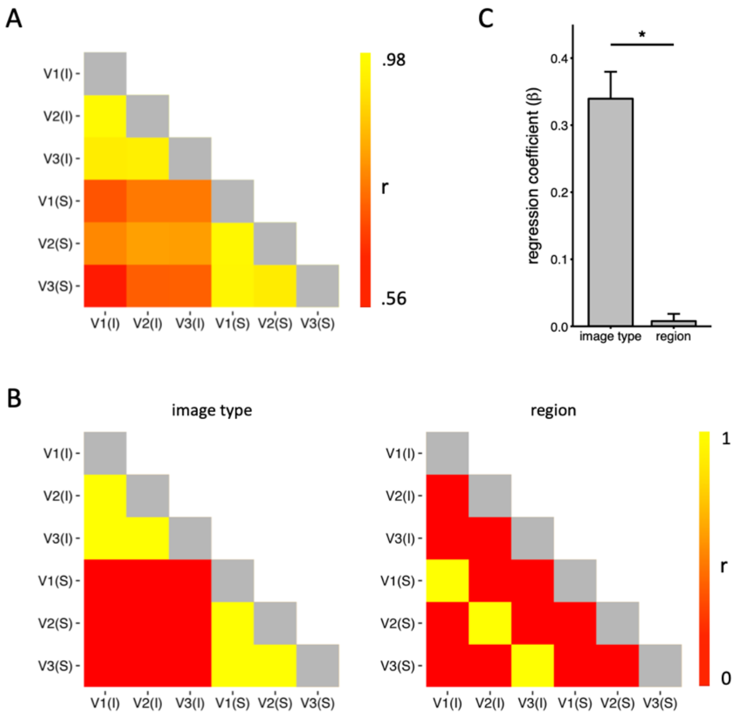

Figure 6.5 *Results of representational similarity analysis. (A) Correlations between the similarity matrices shown in Figure 6.4A. (B) Matrix predictions based on representations of image type and region. (C) Models were used in a regression analysis across participants. Performance was determined by the regression coefficients for each model. The results show that patterns of response were predicted significantly more by the image type than region. Error bars show ±1 SEM.*

** p < .05*

6.5. Discussion

The aim of this study was to determine whether early stages of the ventral visual pathway are selective for objects. To address this issue, we compared both the magnitude and the pattern of response to intact and scrambled images from different object categories in V1, V2 and V3. Our results reveal that all regions showed greater overall neural response to scrambled images relative to intact images. However, this difference was smaller in V3 compared to V1 and V2. We also found that the spatial pattern of response in V3, but not in V1 or V2, was more distinct for intact objects compared to scrambled objects.

The majority of studies of the human object-recognition pathway have focused on the initial (V1) or the final (category-selective regions) stages of processing, while the intermediate stages have received less attention (Peirce, 2015). Neurons in V1 are known to be selective for low-level features of the image (Hubel & Wiesel, 1968). Further downstream, neurons are tuned to properties that appear to combine features encoded in earlier visual areas that are statistically characteristic of natural images (Connor, Brincat, & Pasupathy, 2007; Haxby et al., 2014; Kanwisher, 2010; Tanaka, 1996). Our results show that this selectivity for the properties of natural objects begins to emerge in the response properties of V3. These findings fit with a recent study showing responses to different texture patterns could be differentiated in V3, but not in V2 (Kohler, Clarke, Yakovleva, Liu, & Norcia, 2016).

Scrambled images contain the same visual elements as intact images, but lack the statistical regularities between elements that are important for object perception (Oppenheim & Lim, 1981; Thomson, 1999; Vogels, 1999). Previous studies have shown differential responses to intact and scrambled objects along the ventral visual pathway (Grill-Spector et al., 1998; Malach et al., 1995). Grill-Spector and colleagues used a box-scrambling method to progressively change the degree of scrambling. In early visual areas (V1-V3), they found a higher response to all the scrambled conditions compared to intact images, V4 showed a maximal response to intermediate levels of scrambling and higher visual areas responded most strongly to intact images. These findings suggest that V4 is an important intermediate region in the neural representation of objects (David, Hayden, & Gallant, 2006; Gallant, Braun, & Essen, 1993; Pasupathy & Connor, 2002). However, previous studies have either not reported the responses in V3 or have not distinguished between the response properties of V3 and V1/V2. We found that all regions (V1-V3) showed higher responses to scrambled compared to intact images, but that this difference was attenuated in V3. This suggests that some of the selectivity in V4 to higher order properties of the image may emerge from V3.

The spatial pattern of response to different object categories was more distinct for intact images compared to scrambled images in V3. This analysis is based on a comparison of the within-category similarity in spatial response with the between-category similarity. In a recent study, we found that category-selective patterns of response in high-level regions of the ventral pathway to scrambled images are less distinct than for intact images (Coggan, Liu, et al., 2016). Our current findings suggest that this bias toward natural images begins at an early stage of processing. This reliability in the spatial pattern of response to intact objects is consistent with other studies that have found that temporal patterns of neural response are also more reliable with natural images (Uri Hasson, Malach, & Heeger, 2010).

Although neurons in V2 receive most of their input from V1 and have similar selectivity for orientation and spatial scale (Levitt, Kiper, & Movshon, 1994), a number of studies have shown differences in the response to conjunctions of image features in V2 (Anzai, Peng, & Van Essen, 2007; El-Shamayleh & Movshon, 2011; Ito & Komatsu, 2004). Recent studies have found that neurons in V2 show larger and more reliable responses to synthetic textures that have properties based on natural images compared to control textures (Freeman & Simoncelli, 2011; Freeman et al., 2013; Ziemba et al., 2016). Given this sensitivity to the higher order structure of more naturalistic stimuli, the lack of difference between V1 and V2 in our study was unexpected. One possibility is that the objects used in the current study lack the regularity in structure found in the textures generated by Simoncelli and colleagues (Portilla & Simoncelli, 2000).

Our final analysis involved comparing the representational distances between object categories for intact and scrambled images. Despite the fact that the low-level features were matched between the intact and scrambled images, the representational similarity was more similar for the same image type (intact or scrambled) across regions than for different image types within the same region. For example, the representational similarity between V1 and V2 for intact images was greater that the representational similarity between intact and scrambled images in V1. This suggests that the statistical regularities found in intact images are evident in the pattern of response of early visual areas.

Anatomical observations have shown that neuronal density decreases along the posterior-anterior axis of the primate visual system (Cahalane, Charvet, & Finlay, 2012). This is accompanied by a corresponding reduction in the surface area of regions in higher visual areas (Van Essen et al., 1992). Taken together, these findings indicate that there is a reduction in the amount of information encoded at higher levels of the ventral stream (Lehky, Kiani, Esteky, & Tanaka, 2014). This places constraints on the number of feature conjunctions that can be encoded (Wilson & Wilkinson, 2015). One solution to this combinatorial problem is to only encode combinations of low-level features that are commonly found in natural objects (Kourtzi & Connor, 2011; Kourtzi & Welchman, 2015). Our data shows that the adaptive encoding that is necessary for successful object perception begins at an early stage of processing.

In conclusion, the ventral visual pathway comprises a sequence of cortical areas in which successively more complex visual attributes are extracted, beginning with contour orientations in V1 and resulting in representations of objects at the highest levels. Previous studies have shown that high-level regions of the ventral visual pathway produce greater or more reliable responses to natural, intact images relative to artificial, scrambled images. However, it is currently unclear at which stage in the processing stream this selectivity emerges. Here, we show a preference for natural images can be found at early stages of processing in extrastriate visual cortex.

Chapter 7. The Role of Visual and Semantic Properties in the Emergence of Category-Specific Patterns of Neural Response in the Human Brain

This chapter is adapted from: 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*, ENEURO.0158-16.2016.

7.1. Abstract

Brain imaging studies have found distinct spatial and temporal patterns of response to different object categories across the brain. However, the extent to which these categorical patterns of response reflect higher-level semantic or lower-level visual properties of the stimulus remains unclear. To address this question, we measured patterns of EEG response to intact and scrambled images. Our rationale for using scrambled images is that they have many of the visual properties found in intact images, but do not convey any semantic information. Images from different object categories (bottle, face, house) were briefly presented (400 ms) in an event-related design. A multivariate pattern analysis (MVPA) revealed categorical patterns of response to intact images emerged ~80-100 ms after stimulus onset and were still evident when the stimulus was no longer present (~800 ms). Next, we measured patterns of response to scrambled images. Categorical patterns of response to scrambled images also emerged

~80-100 ms after stimulus onset. However, in contrast to the intact images, distinct patterns of response to scrambled images were mostly evident while the stimulus was present (~400 ms). Moreover, scrambled images were only able to account for all the variance in the intact images at early stages of processing. This direct manipulation of visual and semantic content provides new insights into the temporal dynamics of object perception and the extent to which different stages of processing are dependent on lower-level or higher-level properties of the image.

7.2. Introduction

A full understanding of object perception requires the ability to discriminate object-specific brain states with both spatial *and* temporal resolution. Recently, reliable patterns of neural response to images from different object categories have been shown with MEG and EEG (Carlson et al., 2011, 2013; Cichy et al., 2014; Cauchoix et al., 2014; Clarke et al., 2015). These techniques complement previous MRI studies by providing temporal information about when these categorical patterns of response emerge and how long they are sustained. Temporal properties are important, as they place constraints on models of object recognition (Mur & Kriegeskorte, 2014). Such models suggest a dynamic process in which there is a transformation from a visual representation (based on the statistics of the image) to a semantic representation (reflecting the meaning of the object; Clarke & Tyler, 2015). It is thought that the initial component of the response reflects fast feed-forward processing that is related to visual properties, whereas later patterns reflect recurrent processing that might be related to semantic properties of the stimulus (Bar et al., 2006; DiCarlo & Cox, 2007; Hochstein & Ahissar, 2002; Lamme & Roelfsema, 2000).

The aim of this study was to investigate the relative importance of visual and semantic properties of objects in the emergence of categorical patterns of neural response. However, a fundamental problem in this endeavour is that the visual and semantic properties of objects often covary, making it difficult to resolve the relative contribution of these sources of information to patterns of neural response. So, it is not clear from many previous studies whether the distinct patterns of response to different object categories reflect visual or semantic properties (Carlson et al., 2011, 2013; Cichy et al., 2014; Cauchoix et al., 2014). In a recent MEG study, Clarke and colleagues (2015) addressed this issue by showing that the categorization of objects based on the neural response could be predicted by the visual properties of the image. However, they also found that accuracy could enhanced by including semantic properties, particularly at later stages of processing. Although this suggests that visual and semantic properties are both important for the neural representation of objects, this approach is not able to show a causal link.

To address this issue, we measured patterns of EEG response to intact images from 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 visual properties found in intact images, but they do not convey any semantic information (Coggan, Liu, et al., 2016). This allows us to determine the extent to which the preserved visual properties contribute to the neural representation of objects in the absence of any semantic content. The comparison between the locally scrambled and globally scrambled images also allows us to explore the importance of spatial image properties, which are preserved in the locally scrambled condition. In a recent fMRI study, we found similar spatial patterns of response to intact and scrambled images across the ventral visual pathway (Coggan, Liu, et al., 2016). This study demonstrated the importance of low-level visual properties in the patterns of response in the ventral visual pathway. By comparing the similarity of the responses to intact and scrambled images using EEG, we aim to determine the relative contribution of visual properties to categorical patterns of response at different time-points.

7.3. Methods

7.3.1. Stimuli

105 images of three object categories (face, bottle, house) were taken from an object image stimulus set (Rice et al., 2014). All images were gray-scale, superimposed on a mid-gray background, and had a resolution of 400x400 pixels (Figure 7.1). For each of these original images, 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 8x8 grid and applying the phase-scramble to each 50x50 pixel window independently. In a previous study, we showed that these scrambling significantly attenuates any semantic or categorical content in the images (Coggan, Liu, et al., 2016). Stimuli were presented using a gamma corrected VIEWPixx display (VPixx Technologies Inc., Quebec, Canada) with a resolution of 1920x1200 pixels and a refresh rate of 120Hz. Images were viewed at a distance of approximately 57cm and subtended a retinal angle of 8°.

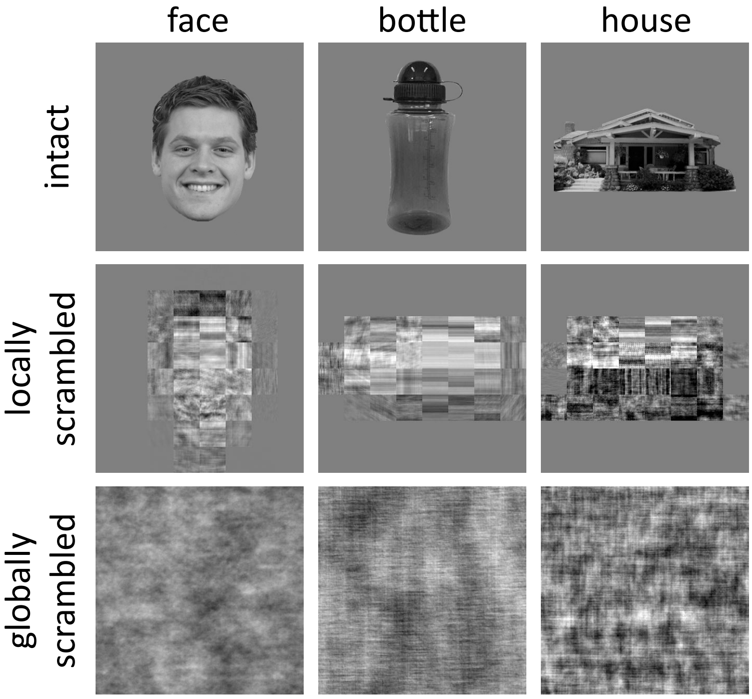

Figure 7.1 *Exemplars of intact, locally scrambled and globally scrambled images from the different object categories.*

7.3.2. Participants

Twenty participants (3 males, mean age = 20.6, SD = 2.6 years) with normal or corrected-to-normal vision took part in the experiment. Participants gave written, informed consent. The study was approved by the University of York Department of Psychology Ethics Committee. The data for one participant (female) was removed from the analysis due to partial data loss.

7.3.3. Design and Procedure

The experiment involved 3 runs: The first run contained globally scrambled images, the second run contained locally scrambled images and the third run contained intact images. Therefore, participants were unaware of the object categories in our stimulus set prior to viewing the scrambled images. Each run contained 35 blocks. There were 10 trials in each block. In each trial, an image from one of the three object categories was presented for 400ms. There was a jittered inter-trial interval that had a mean duration of 1 second and a standard deviation of 200ms. The duration of the inter-block interval was 3 seconds. Participants fixated a cross in the centre of the screen between trials. To maintain attention, participants were instructed to click a mouse whenever a red dot appeared on an image. One image in each block contained a red dot. Self-timed rests were taken between runs.

7.3.4. EEG Recording

EEG waveforms were recorded from 64 scalp locations laid out according to the 10/20 system in a WaveGuard cap (ANT Neuro, Netherlands). Data from each electrode was referenced against a whole-head average. We also monitored blinks through bipolar electrooculogram electrodes placed above and below the left eye. Signals were amplified and digitised at 1000Hz and recorded using the ANT Neuroscan software (ANT Neuro, Netherlands). Stimulus-contingent triggers were sent from the VIEWPixx device to the EEG amplifier using a 25-pin parallel port with microsecond-accurate synchronisation to the display refresh sequence. The PsychToolbox routines (Brainard, 1997; Pelli, 1997) running in Matlab were used to control the display hardware and send triggers.

7.3.5. EEG Pre-processing

The EEG traces from each run were concatenated and band-pass filtered between 0.01-30Hz prior to epoching. Blink artefacts were corrected using independent components analysis (ICA). This involved running ICA across data from all electrodes including the vEOG, and manually selecting the component(s) that captured blink artefacts. These components were then subtracted from the EEG trace at each electrode site according to their weighting. This approach meant that no trials were rejected. The EEG trace was then divided into epochs ranging from 200ms before stimulus onset to 800ms after stimulus onset. All trials containing a red dot were removed prior to further analysis.

7.3.6. EEG MVPA Analysis

All data processing was performed in Matlab using custom scripts. To measure the spatial patterns of EEG response for each participant, trials were collapsed into mean ERPs for odd and even trials for each condition and at each electrode site. These condition-averaged ERPs were then baselined by subtracting the mean amplitude during the 200ms prior to stimulus onset (across both odd and even trials) from the response at each time-point. From these ERPs, a 64-value vector representing the spatial pattern of response across all electrodes was extracted for odd and even trials for each object category at each time-point.

Pattern vectors were normalized within each participant using the following method. First, vectors were selected from one time-point and one image type. This gave a total of 6 patterns (odd/even x face/bottle/house). For each electrode site, the mean amplitude across all 6 patterns was subtracted from its amplitude in each pattern. This process was repeated for each image type at each time-point.

To see whether different object categories evoke distinct patterns of EEG response, we ran a correlation-based MVPA separately for each image type and time-point. This involved measuring the correlation between pattern vectors within and between the three object categories. For within-category correlations (e.g. face vs face), we measured the correlation between odd and even trials. For between-category correlations (e.g. bottle vs house), we used the mean correlation between odd trials of the first category and even trials of the second, and between even trials of the first category and odd trials of the second. The distinctiveness of the patterns of EEG response was then measured by subtracting between-category correlations from within-category correlations. 95% confidence intervals for this difference were then obtained by bootstrapping across participants. Points at which different object categories evoked significantly distinct patterns of EEG response were defined by the lower confidence interval being above zero.

To measure the similarity between responses to intact and scrambled images from the same object category, we first collapsed patterns across odd and even trials to create one pattern per condition per time-point. We then correlated the patterns of response at each time-point separately for the intact-locally scrambled and intact-globally scrambled contrasts for each category. A group mean was calculated across categories and 95% confidence intervals were obtained by bootstrapping across participants.

To determine whether the response to intact images could be explained by the response to scrambled images, we calculated a noise ceiling. This estimates the maximum correlation that could be expected. The noise ceiling was calculated by measuring the correlation between the responses at odd and even trials within each category in the intact condition. At the individual level, we take a mean of the within-category correlations (face-face, bottle-bottle, house-house) for each timepoint. We then average across subjects to obtain one noise ceiling estimate at each timepoint. Timepoints at which this value fell within the 95% CI for the correlation between intact and scrambled images demonstrate when all the variance in the intact images was explained by the scrambled images.

The correlation-based method was complemented with a classification-based approach involving a support vector machine, producing similar results. To see whether different object categories evoked distinct patterns of response, classification was performed separately for each participant, image type and time-point. First, patterns of EEG response were extracted for each trial of each category. Two 'training' patterns and one 'testing' pattern for each category were generated by randomly dividing the 105 trials into three equal sets and taking an average. A support vector machine was then trained on the six training patterns and tested on the three testing patterns. This procedure was repeated 100 times, with different subsets of trials used for training and test in each iteration. To see whether similar patterns of response were evoked by intact and scrambled images from the same category, the classifier was altered so that test patterns were substituted with those from another image type. This was performed for each pairwise contrast between image types, and accuracy was averaged across both directions (e.g. train on intact, test on locally scrambled; and train on locally scrambled, test on intact).

Finally, to examine transient and persistent neural activity in response to each condition, we conducted a temporal cross-correlation. This involved measuring the correlation between response patterns for odd and even trials for the same condition, iterating over each possible pair of time-points. Correlations were represented in a 1000 x 1000 similarity matrix and data were averaged across the positive diagonal. Matrices were then collapsed across categories to give one matrix per image type.

7.4. Results

First, we asked whether different intact object categories produced distinct spatial patterns of EEG response (Figure 7.2). To address this question, we compared the similarity of patterns of response to images from the same category (e.g. face vs face) with the similarity of patterns to images of different categories (e.g. face vs house). Categorical patterns of response were demonstrated when the within-category correlations were significantly greater than the between-category correlations. Categorical patterns of response to intact images emerged 80 ms after stimulus onset. The patterns were maximally distinct at about 150 ms and persisted until at least 800 ms (Figure 7.2B). A classification-based approach was then used to complement the correlation-based method. In this analysis, a classifier was trained on a subset of the data and tested on the remaining data. This showed a similar pattern to the correlation-based analysis. Above chance accuracy emerged 80 ms after stimulus onset, peaked at about 150 ms and persisted until 800 ms (Figure 7.3A).

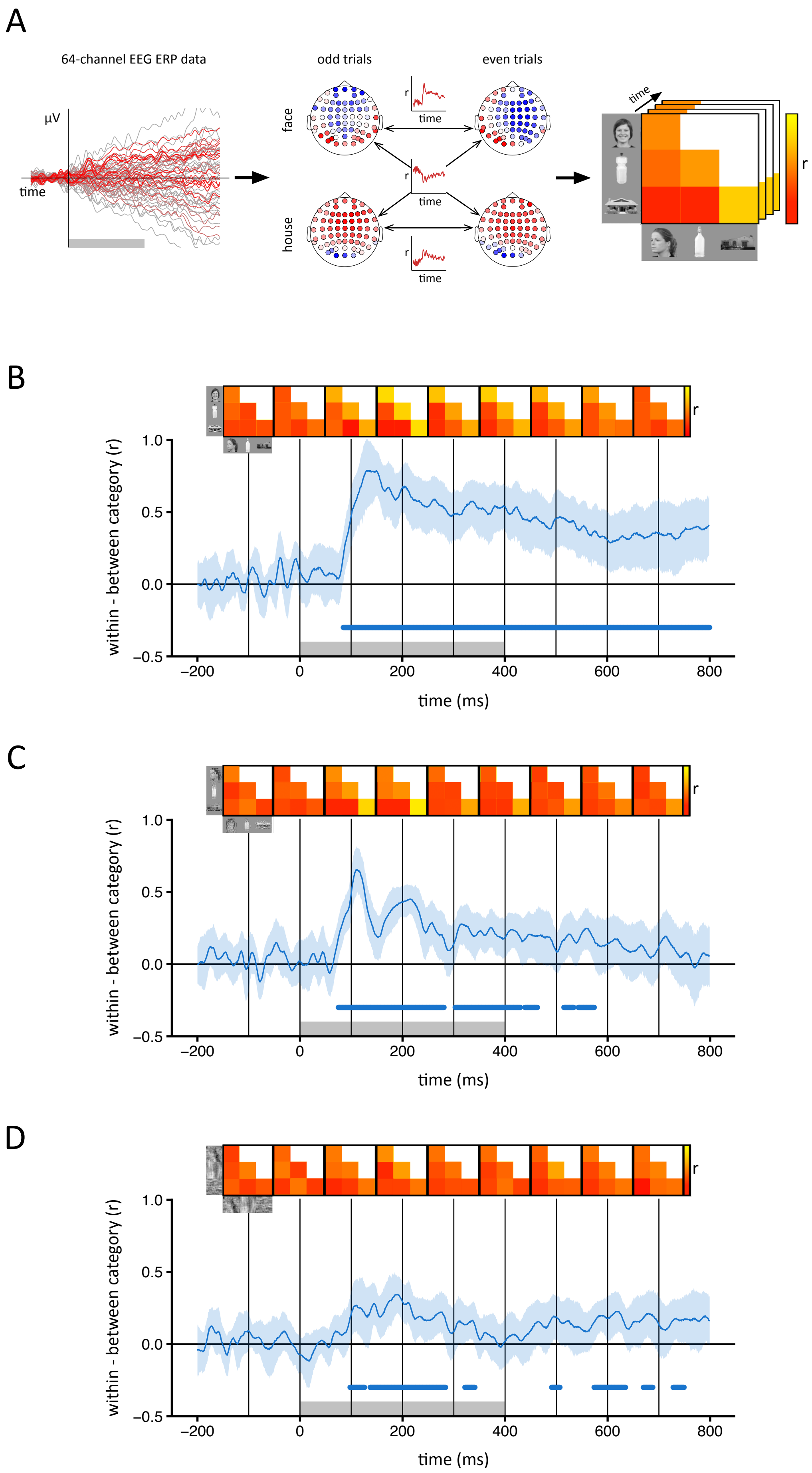
A
64-channel EEG ERP data
µV
time
odd trials
even trials
face
house
r
time
r
time
r
time
time
r
B
C
D
r
within - between category (r)
1.0
0.5
0.0
−0.5
−200
0
200
400
600
800
time (ms)

Figure 7.2 *Category-specific patterns of EEG response to intact and scrambled images. (A) For each time-point, normalized patterns of response to odd and even trials of each category were compared across 64 electrodes. The correlation coefficients were then represented in a similarity matrix for that time-point. Distinct category-specific patterns of response were defined by higher within-category (e.g. face-face) compared to between-category (e.g. face-bottle) correlations. Correlation time-courses are shown for the (B) intact, (C) locally scrambled and (D) globally scrambled image types. Shaded region represents 95% confidence intervals obtained by bootstrapping across participants. Group mean correlation matrices at 100ms intervals are shown above the plot. Grey box at the base of the plot represents the time-points at which the stimulus was present. Blue bar at the base of the plot represents time-points at which the lower bound of the confidence interval is above zero, indicating significantly higher within- than between-category correlations.*

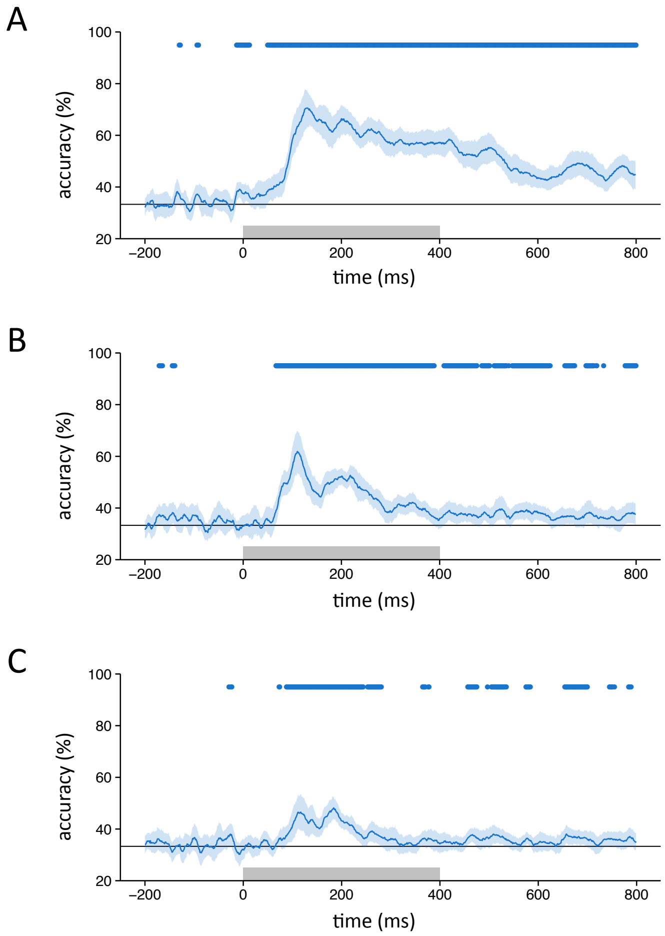

Figure 7.3 *Classifier accuracy for between-category discrimination (blue line) with (A) intact, (B) locally scrambled and (C) globally scrambled images (chance = 33%, grey line). Blue shaded regions represent 95% confidence intervals obtained through bootstrapping across participants. The blue bar at the top of the plot represents time-points at which the lower bound of the confidence interval is above chance. Grey box on the axes of the plot represents the stimulus duration.*

To measure the extent to which these category-specific patterns of response were based on lower-level visual properties, we first asked whether locally scrambled and globally scrambled images also produced distinct category-specific patterns of EEG response using both the correlation-based (Figure 7.2C-D) and classification (Figure 7.3B-C) analyses. Distinct category-specific patterns of response for locally scrambled images emerged after about 80 ms after stimulus onset. They were maximally distinct at about 110 ms and persisted until about 400-500 msec. Distinct category-specific patterns of response for globally scrambled images, emerged at about 100 ms after stimulus onset. They were maximally distinct at about 190 ms and persisted until about 300 msec.

Although distinct patterns of response were evident to scrambled images from different categories (i.e. greater within-category than between-category correlations), it is not clear whether the patterns were similar to those elicited from the intact images. To address this question, we correlated patterns of response to the same object category across different levels of scrambling at different time points. Figure 7.4A (blue horizontal bar) shows that the correlation between intact and locally scrambled images became significant at about 80 ms after stimulus onset, peaked at about 110 ms and 190 msec. The percentage duration that the locally scrambled patterns were correlated with the intact patterns was greater during the stimulus period (0-400 ms: 27%) compared to the post-stimulus period (400-800 ms: 10%). A similar pattern of results was evident when we trained a classifier on intact or locally scrambled images and then tested on locally scrambled or intact images, respectively (Figure 7.5A). The duration of above chance accuracy with the locally scrambled and intact conditions was similar during the stimulus period (0-400 ms: 40%) and the post-stimulus period (400-800 ms: 49%).

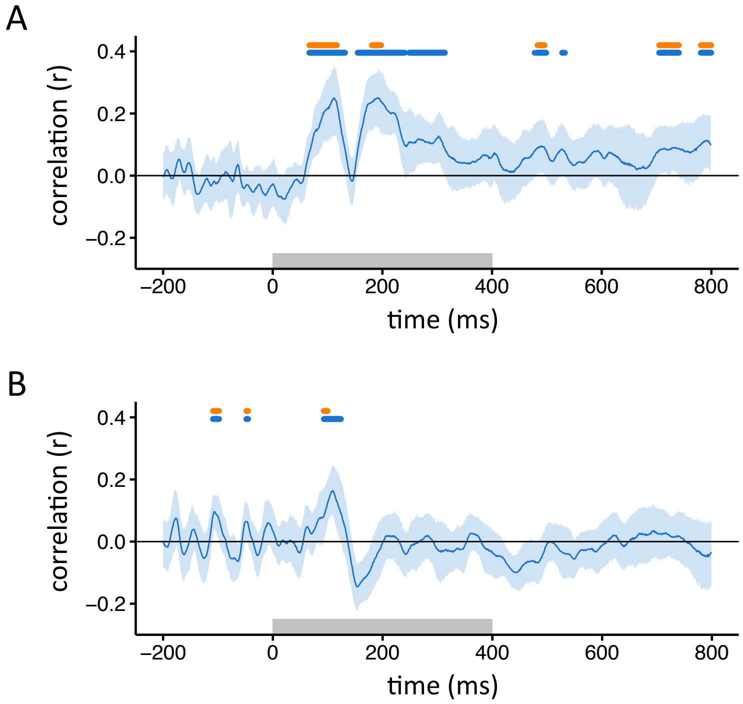

Figure 7.4 *Similarity between patterns of EEG response to intact images and locally scrambled (A) or globally scrambled (B) images from the same object category. Blue shaded regions represent 95% confidence intervals across participants. Blue bar at the top of the plot indicates time-points at which the correlation is significantly above zero. Orange bar indicates time-points at which the correlation is not significantly different from the noise ceiling. Grey box represents the stimulus duration.*

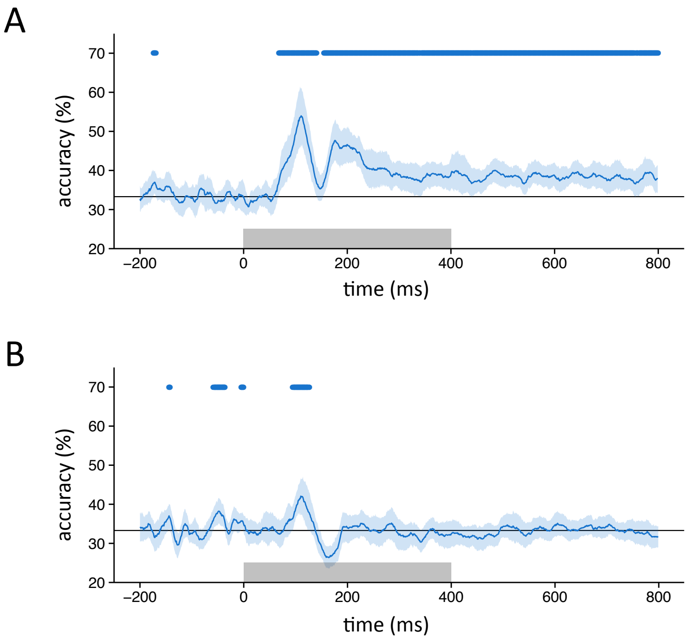

Figure 7.5 *Classifier performance across different image types. (A) Accuracy in classifying responses to either intact or locally scrambled images when trained on locally scrambled or intact images, respectively. (B) Accuracy in classifying responses to either intact or globally scrambled images when trained on globally scrambled or intact images, respectively. Blue line indicates classifier accuracy across time, with shaded regions representing 95% confidence intervals obtained through bootstrapping across participants. Blue bar at the top of the plot represents time-points at which the lower bound of the confidence interval is above chance. Grey box shows stimulus duration.*

Next, we explored the similarity between the intact and globally scrambled images (Figure 7.4B, 7.5B). The correlation between responses to intact and globally scrambled images became significant (blue horizontal bar) about 90 ms after stimulus onset, peaked at about 110 ms and persisted until around 120 msec. The percentage duration that the locally scrambled patterns were correlated with the intact patterns was greater during the stimulus period (0-400 ms: 4%) compared to the post-stimulus period (400-800 ms: 0%). A similar pattern of results was evident when we trained a classifier on intact or locally scrambled images and then tested on locally scrambled or intact images, respectively (Figure 7.5A). The duration of above chance accuracy with the locally scrambled and intact conditions was greater during the stimulus period (0-400 ms: 4%) compared to the post-stimulus period (400-800 ms: 0%).

To directly compare similarity between intact images and either locally scrambled or globally scrambled images, the average correlation (Figure 7.4) or accuracy (Figure 7.5) was compared across individuals. The average correlation between intact and locally scrambled images was significantly higher than the correlation between intact and globally scrambled images (t(18) = 3.29, p <.005). Similarly, the average accuracy (see Figure 7.5) with intact and locally scrambled images was significantly higher than with intact and globally scrambled images (t(18) = 5.34, p < .0001).

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). This involved measuring the correlation to intact images from the same category across odd and even trials of the same category. The noise ceiling was not fixed but varied across time. We then determined whether the correlation between intact and scrambled images was not significantly different to the noise ceiling for each time-point. For locally scrambled images, the 95% confidence intervals of the correlations overlapped until approximately 120 ms after stimulus onset (Figure 7.4A). The percentage duration that the locally scrambled patterns were not significantly different from the noise ceiling was similar during the stimulus (0-400 ms: 9%) and post-stimulus period (400-800 ms: 9%). For globally scrambled images the confidence intervals overlapped until about 100 ms after stimulus onset (Fig 7.4B). The percentage duration that the globally scrambled patterns were not significantly different from the noise ceiling was greater during the stimulus period (0-400 ms: 1%) compared to the post-stimulus period (400-800 ms: 0%).

Finally, we investigated the stability of the category-specific patterns of response for each image manipulation (Cichy et al., 2014). This involved measuring the correlation between patterns of EEG response within each condition across different time-points. The results were then averaged across categories for each image type and represented in time-time similarity matrices (Figure 7.6). Here, the diagonal for intact images corresponds to the noise-ceiling estimate used in Figure 7.4. For intact images, the pattern of response from 100-150 ms was positively correlated with patterns found from ~250-600 msec. The continuation of this neural activity far beyond stimulus offset suggests that this does not reflect prolonged visual input during image presentation. The locally scrambled matrix shows no evidence of persistent neural activity as seen in the intact matrix but does exhibit transient neural activity between ~100-250ms after stimulus onset. Interestingly, time-point combinations of ~150ms and ~200ms show negative correlations, suggesting a polarity reversal in the potentials between these latencies. The globally scrambled matrix shows weak correlations across all combinations of time-points.

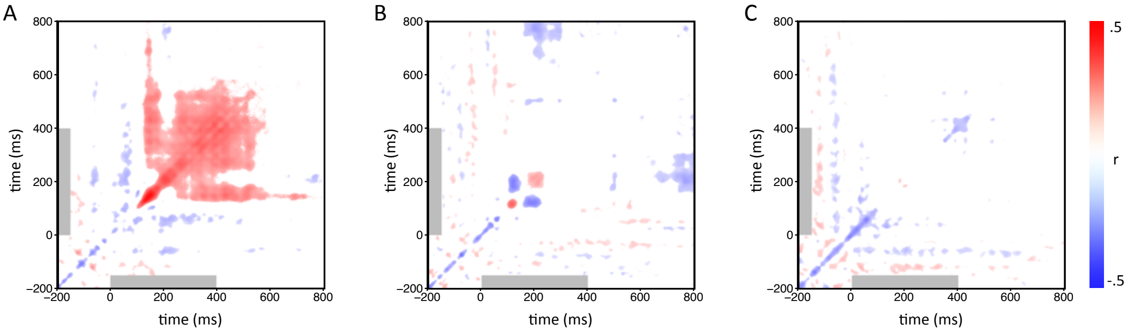

Figure 7.6 *Temporal cross-correlation matrices for each image type. Responses to trials of the same condition were correlated over each combination of time-points. Correlations were collapsed across categories to give one matrix per image type (A - intact, B - locally scrambled, C - globally scrambled). Colourbar represents Pearson's correlation coefficient. Matrices were thresholded by obtaining 95% confidence intervals at each coordinate by bootstrapping across participants. Coordinates at which these intervals overlapped with zero are shown in white. Grey box represents the stimulus duration.*

7.5. Discussion

The aim of this study was to determine the contribution of lower-level visual and higher-level semantic properties to the emergence of categorical patterns of neural response. To address this question, we compared patterns of EEG response to intact and scrambled images from different object categories. Scrambled images were used, because they contain similar visual properties to intact images but do not convey any semantic information (Coggan, Liu, et al., 2016). Our results show similar category-specific patterns of response at early stages of processing. However, these patterns were sustained for a longer time with intact images compared to scrambled images. These results show the importance of visual properties in the emergence of categorical patterns of response, but also show the importance of semantic properties in the recurrent processing that sustains these patterns.

The emergence of category-specific patterns of EEG response to intact images is comparable to previous studies using MEG that found categorical distinctions can be decoded prior to 100ms after stimulus onset and become maximally distinct at about 140 ms (Carlson et al., 2013; Cichy et al., 2014; Cauchoix et al., 2014). However, most previous studies have not directly determined whether these patterns of response reflect lower-level visual properties or higher-level semantic properties of the image. Recently, Clarke and colleagues (2015) addressed this issue with MEG showing that visual properties can explain patterns of response to different categories of objects. However, they also showed that the semantic properties of objects were able to explain additional variance in the pattern of response particularly at later stages of the response. Using scrambled images, which preserve many of the visual properties but do not convey semantic information, we were also able to show that the patterns of response to images from different object categories are driven predominantly by the lower-level visual properties at early stages of visual processing (up to 150 ms). Visual properties were also able to partially account for the variance in the response to intact images at later stages of processing.

Patterns of response to intact images were correlated more strongly and for longer with responses to locally scrambled images than with 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. In Chapters 4, we showed that the spatial pattern of response to different categories of intact objects was more similar to the pattern elicited by locally scrambled objects compared to globally scrambled objects. 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; Golomb and Kanwisher, 2012; Bracci and Op de Beeck, 2015; Silson et al., 2015; Watson et al., 2016).

Although lower-level image properties account for the majority of the variance in responses to intact images at early stages, there remains a significant amount of variance to be explained at later stages of processing. For example, although category-specific patterns of response to intact images persisted well beyond the duration of the stimulus, patterns of response to scrambled images were only evident when the stimulus was present. The persistence of these neural responses to intact images suggests an important role for recurrent processing of the image, which is likely to be driven by top-down semantic representations (Connolly et al., 2012; DiCarlo & Cox, 2007; Kriegeskorte, Mur, Ruff, et al., 2008; Lamme & Roelfsema, 2000; Mur & Kriegeskorte, 2014; Naselaris et al., 2009). Indeed, Clarke and Tyler (2015) showed that accuracy in categorization using MEG data was enhanced by the addition of a semantic model to a visual model.

It is also possible that differences in the patterns of response between intact and scrambled images reflect 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 (Marr & Hildreth, 1980). Indeed, the character and extent of these statistical dependencies is likely to be diagnostic for different classes of images (O'Toole et al., 2005; Rice et al., 2014). The scrambling procedure disrupts many of the statistical relationships between the elements. So, it is possible that image manipulations that can preserve these higher-level visual properties (Freeman & Simoncelli, 2011) might generate responses that are more similar to the intact images. Indeed, it is possible that neural representations underlying higher-level visual properties and the corresponding semantic properties that they convey may be equivalent.

In conclusion, we have found that distinct category-specific patterns of neural response emerge at about 80 ms after stimulus onset and can persist for at least 800 msec. Using scrambled images, we show that early stages of these category-specific patterns can be explained by lower-level image properties. However, the differences in the neural responses to intact and scrambled images at later stages of processing also reveal the importance of higher-level semantic properties.

Chapter 8. General Discussion

Humans are able to categorize the visual scene rapidly and accurately (Potter, 1976; Thorpe et al., 1996). This process of categorization is thought to involve a number of cortical regions, which together constitute the ventral stream (Milner & Goodale, 1995, 2008; Ungerleider & Mishkin, 1982). This ventral stream emerges from the primary visual cortex (area V1), continues through a series of retinotopically organized visual areas (V2, V3, V4) and eventually reaches the ventral temporal cortex, where high-level visual regions are located. Damage to high-level regions can cause various forms of agnosia depending on the location and extent of the lesion (Farah, 1990; McNeil & Warrington, 1993; Mendez & Cherrier, 2003). Further support for the idea that the high-level visual areas have a key role in visual recognition is shown in the selectivity of high-level regions for different categories of objects (Downing et al., 2001; Epstein, 2008; Kanwisher et al., 1997).

Although the selectivity of the ventral stream for objects is clear, the precise functional organisation underlying this selectivity is a matter of current debate. A popular view is that the observed category effects indicate a high-level representation in which neurons are organised around either object category or correlated semantic or conceptual features (Konkle & Oliva, 2012; Kriegeskorte, Mur, Ruff, et al., 2008; Mahon et al., 2009). An alternative view is that categorical responses in the ventral stream are driven by combinations of more basic visual properties that are common to different categories (Andrews et al., 2015; Long, Yu, & Konkle, 2018). The conflation of visual and categorical properties in object images means that category-selective responses could be expected under both accounts.

The aim of this thesis is to investigate the role of basic visual properties in the ventral response to object categories through a number of fMRI and EEG experiments. If the neural representation of objects in the ventral stream is related to visual properties of the image, then groups of objects with different visual properties should elicit distinct patterns of response. In Chapter 3, a data-driven approach was used to select objects based solely on their visual properties. This involved describing the visual properties of a large object database and then using clustering algorithms to objectively select objects from different locations within this feature space. Each cluster of objects had similar visual properties but very different semantic properties. Despite this, each condition evoked a distinct pattern of response across the ventral stream. Additionally, the similarity in response between conditions was strongly predicted by the similarity in visual properties. This suggests that visual properties are an important organising principle in the ventral stream. Further analyses revealed that the neural representation differed across earlier and later regions of the ventral stream. In particular, while responses in early regions were better predicted by visual differences between conditions, responses in later regions were more related to perceptual behaviour. This suggests that later regions are tuned to the key image properties necessary for object perception. Finally, the addition of a noise background to the images reduced the distinctiveness of the response patterns, suggesting that the spatial envelope may be particularly important in the neural response to objects.

If the sensitivity to image properties demonstrated in Chapter 3 can account for category-specific responses in the ventral stream, then such responses should still be evident when images are altered such that categorical information is removed but visual properties are preserved. Chapter 4 describes a second fMRI experiment in which visual and categorical properties were dissociated using image scrambling techniques. Patterns of neural response were measured to images from a range of object categories. Each image was expressed in three versions – one was intact and the other two were scrambled on either a spatially local or global basis. Importantly, a behavioural study showed that human observers could not recognise either type of scrambled images. Yet, distinct patterns of response to different categories were present for intact images and images that had been locally scrambled. Furthermore, categorical patterns of response were similar across these conditions. The lack of semantic properties in the locally scrambled condition constitutes strong evidence that categorical patterns of response in the ventral stream are at least largely driven by the visual properties common to each category. Interestingly, category-specific responses were not evident to globally scrambled images, suggesting that the properties preserved only in the intact and locally-scrambled images are particularly important in driving categorical responses. Finally, responses to locally scrambled images did not fully account for the category-specificity observed for intact images. Therefore, properties disrupted by local scrambling, which include both categorical and visual properties, may also be important in driving category effects in the ventral stream.

Although visual properties can explain a large proportion of the distributed response to objects, they may not account for the response in face- and scene-selective sub-regions of the ventral stream, which exhibit the strongest category-selectivity (Epstein & Kanwisher, 1998; Kanwisher et al., 1997). Chapter 5 describes an fMRI experiment in which responses to intact and scrambled images of faces and houses were measured in the fusiform face area and parahippocampal place area. Both regions showed significantly greater response to their preferred than non-preferred category for each of the intact, locally scrambled and globally scrambled conditions. Both regions also showed greater adaptation for the preferred category compared to the non-preferred category in both intact and locally scrambled conditions. Again, the lack of categorical information in the locally scrambled images provides strong evidence for the role of visual properties in the responses of even the most category-selective regions of the ventral stream. Also, as in Chapter 4, category effects produced by locally scrambled images were larger than those by globally scrambled images, but smaller than those by intact images. This reinforces the role of both spatial properties and properties disrupted by local scrambling in category-selective responses. We also observed significant, albeit smaller, adaptation for intact, non-preferred categories. This suggests that the representation of faces and scenes is distributed across the ventral surface rather than localised within discrete cortical regions.

A secondary finding in Chapter 4 was that early visual regions showed greater response to scrambled object images over natural (intact) images. However, Chapter 5 and previous studies (Andrews et al., 2010; Grill-Spector, Kushnir, Hendler, & Malach, 2000; Malach et al., 1995) showed that later regions of ventral stream exhibit the opposite preference. Chapter 6 presents an fMRI study investigating where along the ventral visual pathway the preference for natural object images emerges. This was achieved by comparing responses in early visual regions (V1,V2,V3) to intact and locally scrambled images across a range of categories. Regions V1 and V2 responded similarly - both showed a similar preference for scrambled images and similar-strength categorical patterns of responses for intact and scrambled images. However, V3 showed a reduced preference for scrambled images, and stronger categorical patterns for intact versus scrambled images. These results show that the selectivity for natural images observed in later regions of the ventral stream emerges at an early stage of processing. This selectivity could represent adaptive encoding in which particular combinations of low-level features that are common in natural images and / or useful for object discrimination become increasingly represented as one moves up the ventral visual hierarchy. These relationships are disrupted in scrambled images, explaining why they produced weaker responses. Lastly, we observed that the representation of the categories was highly similar across regions, but different across intact and scrambled images. This suggests that the statistical regularities found in intact images are evident in the pattern of response in early visual areas.

A comprehensive understanding of visual object perception requires the neural response to be characterised in space and time. Recent studies have found category effects emerging shortly after stimulus onset (<100 ms) which, if reflective of categorical representations, imposes strict temporal constraints on models of object recognition such that faster, feedforward models are preferred over recurrent processing models (Mur & Kriegeskorte, 2014). However, as in the fMRI literature, there is uncertainty regarding the relative influence of visual and categorical properties in these responses. To address this, Chapter 7 presents an EEG experiment in which patterns of response across the scalp were measured to intact, locally scrambled and globally scrambled images across a range of object categories. In line with previous studies, intact images produced category effects from around 80 ms after stimulus onset, which persisted even when the stimulus was no longer present. However, these effects were fully replicated using locally scrambled images until approximately 120 ms after stimulus onset, suggesting that visual properties are entirely responsible for category effects up to this point. Between 120 ms and stimulus offset, there was partial correspondence between responses to intact and locally scrambled images, suggesting that visual properties may only partially account for categorical responses during this stage. After stimulus offset, only intact images reliably produced category effects, indicating that categorical information is important in the recurrent processing that sustains the patterns of response beyond the stimulus duration. This provides new information about the time-course of visual object processing, and suggests that high-level object representations may take longer to construct (at least 120 ms) than recent estimates have suggested (Carlson et al., 2011; Carlson et al., 2013; Cichy et al., 2014). As in previous chapters, responses to intact images were better predicted by responses to locally scrambled images than globally scrambled images. Thus, the role of spatial properties in visual object processing is evident in both electrophysiological and BOLD responses.

Thus, the experiments presented in this thesis provide evidence that the neural response in the ventral stream is driven largely by more basic visual properties of the image. This is consistent with previous studies that have shown that visual differences between object categories predict differences in subsequent neural responses (O'Toole et al., 2005; Rice et al., 2014; Watson et al., 2014; Watson, Hymers, et al., 2016). However, these previous studies were correlational in nature and, due to the link between visual and categorical properties in object images, the prediction of categorical patterns by visual descriptions is expected under both high- and low-level accounts. This thesis builds on previous work by forging a dissociation between visual and categorical properties, finding across several studies that a substantial proportion of category effects on the ventral response can be directly attributed to more basic visual representations.

Which visual properties are important in driving categorical responses to objects? A common finding across Chapters 4,5 and 7 was that locally scrambled images perform significantly better than globally scrambled images in producing category effects. There are several differences between these two scrambling procedures regarding which visual properties are preserved. Specifically, local scrambling coarsely preserves the spatial envelope (i.e. the space within the image taken up by the object) which is also strongly linked to its overall shape. The retinotopy / spatial position of orientation and spatial frequency information is also largely retained. However, these spatial properties are entirely destroyed by global scrambling. Spatial properties therefore likely constitute the key visual features responsible for driving category effects in the ventral stream, although global properties did show some success. However, while spatial properties are entirely visual, this term captures both mid-level (e.g. shape) and low-level (e.g. retinotopy) features. This raises the issue of precisely which spatial properties are important. Chapter 3 yielded two findings relevant to this question. First, the neural representation in later regions of the ventral stream was different to early visual regions, although both were related to visual properties. This could suggest that categorical responses cannot be entirely reduced to the distribution of orientation and spatial frequency information across the image space, as these features strongly characterise the response tuning of early visual cortex (Bonhoeffer & Grinvald, 1991; Engel et al., 1994; Hubel & Wiesel, 1968; Wandell et al., 2007). Second, the addition of a noise background, which diminishes the spatial envelope of the object, substantially reduced the distinctiveness of condition-specific responses. The spatial envelope and related properties, such as overall shape, may therefore be particularly relevant. This is in line with recent studies showing the importance of shape information in the ventral response to categories (Bracci & Op de Beeck, 2016; Long et al., 2018; Proklova et al., 2016). Taken together, this suggests that higher-order visual cues such as shape may be particularly important in generating category effects in the ventral stream.

Another common finding across the experiments presented in this thesis is that the strength of category effects found for intact images was not matched by scrambled images. So, while visual properties may explain a substantial amount of the observed category-selectivity, there is residual selectivity that remains unaccounted for. This could be explained by a tuning to visual features that are not preserved even by local scrambling but are nevertheless correlated with object category. For instance, local scrambling destroys textural information; smaller-scale shape cues within the object and the fine-scale spatial envelope. In line with this view, recent studies have found that texture and shape cues of different categories are sufficient to predict and produce categorical responses to objects in the ventral stream. For example, one study found that the difference between patterns of ventral response to a range of objects was predicted by the similarity in both their outline and their texture, measured behaviourally (Proklova et al., 2016). Another study found that object images scrambled in a way that preserved shape and textural cues produced similar patterns of ventral response to the original intact images (Long et al., 2018). An alternative explanation for the residual category-selectivity we observed is that high-level organising principles form part of the neural representation in this region. In support of this view, recent investigations have involved great efforts to disentangle visual and categorical properties experimentally (Bracci & Op de Beeck, 2016; Proklova et al., 2016). In these studies, stimuli were selected such that objects from the same category each had a different set of visual properties and objects with a similar set of visual properties each belonged to different a category. Despite this, categorical effects were still present, suggesting that category effects cannot be entirely reduced to shape and texture cues. Ultimately, we cannot know the extent to which the residual category effects we observed is a product of categorical or visual representations.

It is important to note that the findings presented in this thesis do not dispute that the ventral visual pathway is selective for categories. Instead, they provide evidence for a particular mechanism by which such selectivity might arise. Rather than encoding high-level properties such as category membership or conceptual attributes, it may be that they represent different combinations of visual properties that are common to different categories. This is consistent with a previously proposed model of how category-selectivity arises in the ventral stream (Op de Beeck et al., 2008). In this model, a number of topographical maps are superimposed on the same expanse of cortex, each one weakly selective for a particular visual feature, for instance orientation, curvature or shape. Importantly, these maps are spatially correlated with one another such that the visual properties common to a particular object category are represented in similar locations across the maps. Although each map has weak response bias, the responses are combined across maps in a multiplicative fashion in order to generate a large, localized increase in neural activation. This could theoretically give rise to a strongly category-selective region that shows weak response biases for a number of visual properties. A simplified illustration is shown in Figure 8.1. It should also be noted that these maps need not be spatially correlated with one another in order to account for category-specific responses. A similar model proposed by Andrews et al. (2015) shows that even entirely orthogonal visual maps superimposed upon one another could give rise to category-specific patterns of response, provided that different categories have different characteristic visual properties.

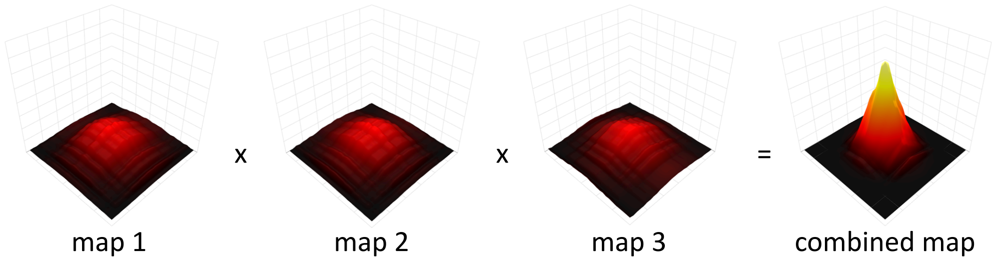

Figure 8.1 *Illustration of Op de Beeck et al.'s (2008) model of how category selectivity arises. A number of maps are superimposed on the same patch of cortex, each weakly tuned to a particular visual feature. However, the maps are spatially correlated such that a particular category will activate similar regions of each map. The responses across maps are combined multiplicatively in order to produce a strong, localised response to a particular category.*

In addition to explaining both distributed and localised responses to object categories, the models proposed by Op de Beeck et al. (2008) and Andrews et al. (2015) offer a number of advantages over traditional accounts of the neural organisation, in which neurons respond according to semantic or conceptual properties of the stimulus, such as its category. First, it is not clear how such a high-level organisation emerges from the image-based representation found in preceding visual regions. These visual models suggest that no explanation is necessary – while later regions likely process higher-order combinations of low-level properties, the representation is nevertheless fundamentally visual. Second, it is difficult to reconcile a categorical organisation with previous findings of response biases for visual properties such as spatial frequency, curvature, shape and retinotopy (Arcaro et al., 2009; Brewer et al., 2005; Hasson, Levy, Behrmann, Hendler, & Malach, 2002; Levy et al., 2001; Nasr et al., 2014; Rajimehr et al., 2011; Silson et al., 2015; Weiner et al., 2014). These biases have been previously dismissed as by-products of a tuning to more complex, high-level features, and suggested to be too weak to account for the strong effects of category (Op de Beeck et al., 2008). However, the visual models place these biases in a direct functional role. Furthermore, the weakness of individual biases does not limit the explanatory power of visual properties as an organizing principle. Rather, these individual biases are combined multiplicatively in order to support strong, localised category-selectivity. Finally, the visual models are entirely consistent with the observations in this thesis that a substantial proportion of the category-selectivity is still present when images are scrambled beyond recognition. In particular, combinations of visual response biases (e.g. retinotopy, spatial frequency, orientation) are preserved in the locally scrambled stimuli. So, locally scrambled stimuli may have produced stronger categorical responses than globally scrambled images by virtue of preserving the features necessary for the multiplicative combination of responses described by the model.

An advantage of a fundamentally image-based representation in high-level visual cortex is that it can be used more flexibly in the processing of objects. Although object recognition appears to be a central goal of the ventral stream, neural responses have also been linked to navigation (Epstein, 2008), social perception (Downing & Peelen, 2011), facial identification (Andrews & Ewbank, 2004) and reading (Cohen et al., 2000). These different tasks likely require different forms of visual analysis, for instance face perception appears to involve more holistic processing than the perception of other categories (McKone, Kanwisher, & Duchaine, 2007; Robbins & McKone, 2007). The diversity in visual tasks associated with the ventral stream has been interpreted in favour of an organisation in which category-selective regions each encode the stimulus properties relevant to the particular range of perceptual goals associated with its preferred category (Downing & Peelen, 2017). This is consistent with the notion that each category-selective region represents an intersection of relevant visual properties that can be selectively extracted according to task demands.

The selective extraction or enhancement of information from the ventral stream is supported by findings that ventral responses are modulated by attention (Kay, Weiner, & Grill-Spector, 2015), task (Harel, Kravitz, & Baker, 2014) and imagery (O’Craven & Kanwisher, 2000). Importantly, these studies do not by themselves constitute evidence of high-level representations in the ventral stream, as similar effects can be found in early visual cortex (Hsieh, Vul, & Kanwisher, 2010; Slotnick, Thompson, & Kosslyn, 2005). Instead, these effects may be the result of top-down modulation that allows different visual features to be extracted from the ventral stream according to task demands. Such feedback has been suggested to occur during object recognition, effects of which have been detected in orbito-frontal cortex 50 ms before they emerge in the ventral stream (Bar et al., 2006). Aside from top-down feedback, it is also possible that responses are modulated by lateral connections, perhaps in the form of cross-modal input (Wolbers, Klatzky, Loomis, Wutte, & Giudice, 2011). A fundamentally visual organisation provides a task-neutral representation, albeit biased toward properties that are critical for perception, from which information relevant to the current perceptual goals can be extracted.

It has been suggested that distributed patterns of response in the ventral stream are largely idiosyncratic to each individual subject (Haxby et al., 2014). In this study, patterns of response were found to be more consistent across different subjects when voxels were ‘hyperaligned’ - a technique whereby voxels are matched across subjects based on the similarity in response across a range of visual stimuli - as opposed to the traditional approach in which voxels are aligned by their anatomical location. This suggests a degree of individual difference in the precise topography of the ventral stream. However, others have noted that category-selective regions are found in approximately the same location in different subjects (Kanwisher, 2010) and a number of studies have successfully cross-validated patterns of response across participants (Poldrack et al., 2009; Rice et al., 2014; Shinkareva et al., 2008). The fMRI studies presented in Chapters 3 and 4 of this thesis also employed a between-subjects cross-validation paradigm. The findings reported here therefore support the notion of organising principles that are generally consistent across the population.

In conclusion, this thesis aimed to measure the role of basic visual features in category-selective neural responses in the ventral visual pathway. A series of fMRI and EEG experiments revealed that distinct responses are still present when images are grouped by image properties rather than category. Furthermore, category-effects were still present when images were scrambled beyond recognition, even in highly category-selective regions. This suggests that a tuning to more basic visual properties underlies the category-selectivity observed in the ventral stream. These and previous findings can be reconciled by a neural organisation in which an array of different lower-level feature maps are superimposed on the same expanse of cortex. Taken together, these results contribute to the literature by characterising the basic visual dimensions underlying the neural representation of object categories in space and time.

Appendices

A.1 Supplementary Figures

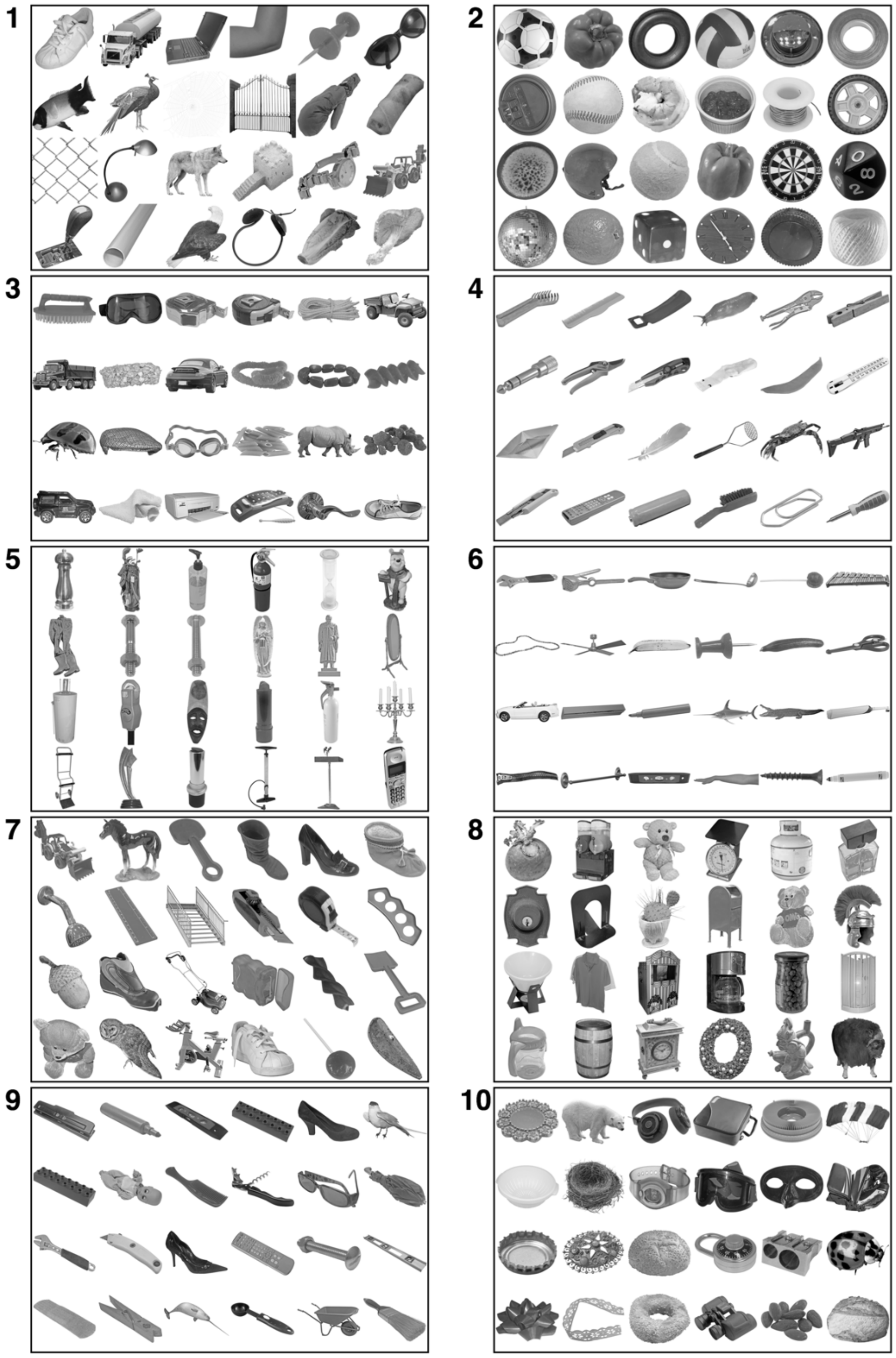

Figure A.1.1 *Full object stimulus set for Chapter 3 prior to application of backgrounds.*

A.2 Supplementary Tables

		Effect of Comparison (F)			Correlation between similarity matrices (*r*)	
Mask	Overlap retinotopic regions (%)	Intact	Locally-scrambled	Globally-scrambled	Intact v Locally-scrambled	Intact v Globally-scrambled
ventral stream	28	259.91***	248.16***	25.00***	.88***	.53*
ventral stream (w/o V1-V4)	10	261.93***	152.96***	23.02***	.86***	.55*
ventral stream (w/o V1-V4, VO1/2, PH1/2)	0	276.02***	143.24***	22.92***	.86***	.58*
posterior-anterior 1	59	129.14***	105.54***	24.36***	.86***	.55*
posterior-anterior 2	47	161.15***	163.70***	12.12**	.83***	.34
posterior-anterior 3	13	214.08***	29.95***	8.47*	.92***	.54*
posterior-anterior 4	11	169.90***	28.52***	5.32*	.62*	.44

Table A.2.1 *Summary of analyses with different masks. The effect of comparison was calculated by determining the difference between within-category versus between-category correlations. The extent to which intact and scrambled images from the same category evoked similar patterns of response is shown in the correlation between similarity matrices. * p < .05, ** p < .005, *** p < .0005*

References

Allison, T., Ginter, H., McCarthy, G., Nobre, A., Puce, A., Luby, M., & Spencer, D. (1994). Face recognition in human extrastriate cortex. *Journal of Neurophysiology*, *71*, 821–825.

Andrews, T. J., Baseler, H., Jenkins, R., Burton, A. M., & Young, A. W. (2016). Contributions of feature shapes and surface cues to the recognition and neural representation of facial identity. *Cortex*, *83*, 280–291.

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., & Ewbank, M. P. (2004). Distinct representations for facial identity and changeable aspects of faces in the human temporal lobe. *NeuroImage*, *23*(3), 905–913.

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.

Anzai, A., Peng, X., & Van Essen, D. C. (2007). Neurons in monkey visual area V2 encode combinations of orientations. *Nature Neuroscience*, *10*(10), 1313–1321.

Arcaro, M. J., McMains, S. A., Singer, B. D., & Kastner, S. (2009). Retinotopic organization of human ventral visual cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *29*(34), 10638–10652.

Avidan, G., Hasson, U., Hendler, T., Zohary, E., & Malach, R. (2002). Analysis of the Neuronal Selectivity Underlying Low fMRI Signals. *Current Biology*, *12*(12), 964–972.

Bar, M., Kassam, K., Ghuman, A., Boshyan, J., Schmidt, A., Dale, A., … Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(2), 449–454.

Beauchamp, M. S., Lee, K. E., Haxby, J. V., & Martin, A. (2002). Parallel visual motion processing streams for manipulable objects and human movements. *Neuron*, *34*(1), 149–159.

Behrmann, M., & Plaut, D. C. (2013). Distributed circuits, not circumscribed centers, mediate visual recognition. *Trends in Cognitive Sciences*, *17*(5), 210–219.

Benjamini, Y., & Hochberg, Y. (1995). Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B*, *57*(1), 289–300.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review*, *94*(2), 115–147.

Bonhoeffer, T., & Grinvald, A. (1991). Iso-orientation domains in cat visual cortex are arranged in pinwheel-like patterns. *Nature*, *353*(6343), 429–431.

Bouvier, S. E., & Engel, S. A. (2006). Behavioral deficits and cortical damage loci in cerebral achromatopsia. *Cerebral Cortex*, *16*(2), 183–191.

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.

Bracci, S., Ritchie, J. B., & Op de Beeck, H. (2017). On the partnership between neural representations of object categories and visual features in the ventral visual pathway. *Neuropsychologia*, (October 2016), 0–1.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1017/CBO9781107415324.004

Brewer, A. a, Liu, J., Wade, A. R., & Wandell, B. a. (2005). Visual field maps and stimulus selectivity in human ventral occipital cortex. *Nature Neuroscience*, *8*(8), 1102–1109.

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).

Cahalane, D. J., Charvet, C. J., & Finlay, B. L. (2012). Systematic, balancing gradients in neuron density and number across the primate isocortex. *Frontiers in Neuroanatomy*, *6*, 28.

Cant, J. S., & Xu, Y. (2012). Object Ensemble Processing in Human Anterior-Medial Ventral Visual Cortex. *Journal of Neuroscience*, *32*(22), 7685–7700.

Carlson, T. A., Hogendoorn, H., Kanai, R., Mesik, J., & Turret, J. (2011). High temporal resolution decoding of object position and category. *Journal of Vision*, *11*(2011), 1–17.

Carlson, T. a, Schrater, P., & He, S. (2003). Patterns of activity in the categorical representations of objects. *Journal of Cognitive Neuroscience*, *15*(5), 704–717.

Carlson, T., Tovar, D., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics of object vision: The first 1000 ms. *Journal of Vision*, *13*(10), 1–19.

Cauchoix, M., Barragan-Jason, G., Serre, T., & Barbeau, E. J. (2014). The neural dynamics of face detection in the wild revealed by MVPA. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *34*(3), 846–854.

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.

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience*, *17*(3), 455–462.

Cichy, R. M., Sterzer, P., Heinzle, J., Elliott, L. T., Ramirez, F., & Haynes, J. D. (2013). Probing principles of large-scale object representation: Category preference and location encoding. *Human Brain Mapping*, *34*(7), 1636–1651.

Clark, V. P., Keil, K., Maisog, J. M., Courtney, S., Ungerleider, L. G., & Haxby, J. V. (1996). Functional Magnetic Resonance Imaging of Human Visual Cortex during Face Matching: A Comparison with Positron Emission Tomography. *NeuroImage*, *4*, 1–15.

Clarke, A., Devereux, B. J., Randall, B., & Tyler, L. K. (2015). Predicting the time course of individual objects with MEG. *Cerebral Cortex*, *25*(10), 3602–3612.

Clarke, A., & Tyler, L. (2014). Object-Specific Semantic Coding in Human Perirhinal Cortex. *The Journal of Neuroscience*, *34*(14), 4766–4775.

Clarke, A., & Tyler, L. K. (2015). Understanding What We See: How We Derive Meaning

From Vision. *Trends in Cognitive Sciences*, *19*(11), 677–687.

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*(August), 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., Watson, D. M., Hartley, T., Baker, D. H., & Andrews, T. J. (under review). A data-driven approach to stimulus selection reveals the emergence of an image-based representation of objects in high-level visual areas. *Cerebral Cortex.*

Coggan, D. D., Baker, D. H., & Andrews, T. J. (in press). Selectivity for mid-level properties of faces and places in the fusiform face area and parahippocampal place area. *European Journal of Neuroscience.*

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.

Connor, C. E., Brincat, S. L., & Pasupathy, A. (2007). Transformation of shape information in the ventral pathway. *Current Opinion in Neurobiology*, *17*(2), 140–147.

Cox, D. D., & Savoy, R. L. (2003). Functional magnetic resonance imaging (fMRI) “brain reading”: detecting and classifying distributed patterns of fMRI activity in human visual cortex. *NeuroImage*, *19*(2), 261–270.

Dakin, S. C., & Watt, R. J. (2009). Biological “bar codes” in human faces Steven. *Journal of Vision*, *9*(2009), 1–10. https://doi.org/10.1167/9.4.2.Introduction

David, S. V, Hayden, B. Y., & Gallant, J. L. (2006). Spectral receptive field properties explain shape selectivity in area V4. *Journal of Neurophysiology*, *96*(6), 3492–3505.

De Renzi, E., Perani, D., Carlesimo, G. a, Silveri, M. C., & Fazio, F. (1994). Prosopagnosia can be associated with damage confined to the right hemisphere--an MRI and PET study and a review of the literature. *Neuropsychologia*, *32*(8), 893–902.

Deselaers, T., & Ferrari, V. (2011). Visual and semantic similarity in ImageNet. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 1777–1784).

DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in Cognitive Sciences*, *11*(8), 333–341.

Downing, P. E., Chan, A. W.-Y., Peelen, M., Dodds, C. M., & Kanwisher, N. (2006). Domain Specificity in Visual Cortex. *Cerebral Cortex*, *16*, 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.

Downing, P. E., & Peelen, M. V. (2011). The role of occipitotemporal body-selective regions in person perception. *Cognitive Neuroscience*, *2*(3–4), 186–203.

Downing, P. E., & Peelen, M. V. (2017). Category selectivity in human visual cortex: Beyond visual object recognition. *Neuropsychologia*, *105*(March), 177–183.

Drucker, D. M., & Aguirre, G. K. (2009). Different spatial scales of shape similarity representation in lateral and ventral LOC. *Cerebral Cortex*, *19*(10), 2269–2280.

Dumoulin, S. O., & Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. *NeuroImage*, *39*(2), 647–660.

Edelman, S. (1998). Representation Is Representation of Similarities. *Behavioral and Brain Sciences*, *21*, 449–498.

Edelman, S., Grill-Spector, K., Kushnir, T., & Malach, R. (1998). Toward direct visualization of the internal shape representation space by fMRI. *Psychobiology*, *26*(4), 309.

Eger, E., Schyns, P. G., & Kleinschmidt, A. (2004). Scale invariant adaptation in fusiform face-responsive regions. *NeuroImage*, *22*(1), 232–242.

El-Shamayleh, Y., & Movshon, J. A. (2011). Neuronal Responses to Texture-Defined Form in Macaque Visual Area V2. *Journal of Neuroscience*, *31*(23), 8543–8555.

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. A. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends in Cognitive Sciences*, *12*(10), 388–396.

Epstein, R., Graham, K. S., & Downing, P. E. (2003). Viewpoint-specific scene representations in human parahippocampal cortex. *Neuron*, *37*(5), 865–876.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*(6676), 598–601.

Ewbank, M. P., Schluppeck, D., & Andrews, T. J. (2005). fMR-adaptation reveals a distributed representation of inanimate objects and places in human visual cortex. *NeuroImage*, *28*(1), 268–279.

Farah, M. J. (1990). *Visual Agnosia*. Cambridge, MA: MIT Press.

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.

Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nature Neuroscience*, *14*(9), 1195–1201.

Freeman, J., Ziemba, C. M., Heeger, D. J., Simoncelli, E. P., & Movshon, J. A. (2013). A functional and perceptual signature of the second visual area in primates. *Nature Neuroscience*, *16*(7), 974–981.

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. -P, Frith, C. D., & Frackowiak, R. S. J. (1994). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, *2*(4), 189–210.

Gallant, J. L., Braun, J., & Essen, D. C. Van. (1993). Selectivity for Polar , Hyperbolic , and

Cartesian Gratings in Macaque Visual Cortex. *Science*, *259*, 100–103.

Gauthier, I., Tarr, M. J., Moylan, J., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). The fusiform “face area” is part of a network that processes faces at the individual level. *Journal of Cognitive Neuroscience*, *12*(3), 495–504.

Geisler, W. S. (2008). Visual perception and the statistical properties of natural scenes. *Annual Review of Psychology*, *59*, 167–192.

Goffaux, V., & Dakin, S. C. (2010). Horizontal information drives the behavioral signatures of face processing. *Frontiers in Psychology*, *1*(SEP), 1–14.

Goffaux, V., Duecker, F., Hausfeld, L., Schiltz, C., & Goebel, R. (2016). Horizontal tuning for faces originates in high-level Fusiform Face Area. *Neuropsychologia*, *81*(2016), 1–11.

Golomb, J. D., & Kanwisher, N. (2012). Higher level visual cortex represents retinotopic, not spatiotopic, object location. *Cerebral Cortex*, *22*(12), 2794–2810.

Grill-Spector, K., & Kanwisher, N. (2005). Visual Recognition: As soon as you know it is there, you know what it is. *Psychological Science*, *16*(2), 152–160.

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., & Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron*, *24*(1), 187–203.

Grill-Spector, K., Kushnir, T., Edelman, S., Itzchak, Y., & Malach, R. (1998). Cue-invariant activation in object-related areas of the human occipital lobe. *Neuron*, *21*(1), 191–202.

Grill-Spector, K., Kushnir, T., Hendler, T., & Malach, R. (2000). The dynamics of object-selective activation correlate with recognition performance in humans. *Nature Neuroscience*, *3*(8), 837–843.

Grill-Spector, K., & Malach, R. (2001). fMR-adaptation: A tool for studying the functional properties of human cortical neurons. *Acta Psychologica*, *107*(1–3), 293–321.

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.

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.

Harel, A., Kravitz, D. J., & Baker, C. I. (2014). Task context impacts visual object processing differentially across the cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *10*, 85–86.

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*(3), 479–490.

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*(3), 479–490.

Hasson, U., Malach, R., & Heeger, D. J. (2010). Reliability of cortical activity during natural stimulation. *Trends in Cognitive Sciences*, *14*(1), 40–48.

Haushofer, J., Livingstone, M. S., & Kanwisher, N. (2008). Multivariate patterns in object-selective cortex dissociate perceptual and physical shape similarity. *PLoS Biology*, *6*(7), 1459–1467.

Haxby, J. V., Connolly, A. C., & Guntupalli, J. S. (2014). Decoding Neural Representational Spaces Using Multivariate Pattern Analysis. *Annual Review of Neuroscience*, *37*, 435–456.

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.

Haynes, J.-D. (2015). A Primer on Pattern-Based Approaches to fMRI: Principles, Pitfalls, and Perspectives. *Neuron*, *87*(2), 257–270.

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, *36*(5), 791–804.

Hodges, J. R., Patterson, K., Oxbury, S., & Funnell, E. (1992). Semantic dementia: progressive fluent apahsia with temporal lobe atrophy. *Brain*, *115*, 1783–1806.

Honey, C., Kirchner, H., & VanRullen, R. (2008). Faces in the cloud: Fourier power spectrum biases ultrarapid face detection. *Journal of Vision*, *8*(2008), 9.1-13.

Hsieh, P.-J., Vul, E., & Kanwisher, N. (2010). Recognition alters the spatial pattern of FMRI activation in early retinotopic cortex. *Journal of Neurophysiology*, *103*(3), 1501–1507.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive Fields and Functional Architecture of monkey striate cortex. *Journal of Physiology*, *195*, 215–243.

Humphries, C., Liebenthal, E., & Binder, J. (2010). Tonotopic organization of human auditory cortex. *NeuroImage*, *50*(3), 1202–1211.

Hung, C. P., Kreiman, G., Poggio, T., & DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science (New York, N.Y.)*, *310*(5749), 863–866.

Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(16), 9379–9384.

Ito, M., & Komatsu, H. (2004). Representation of Angles Embedded within Contour Stimuli in Area V2 of Macaque Monkeys. *Journal of Neuroscience*, *24*(13), 3313–3324.

Jenkins, R., White, D., Van Montfort, X., & Burton, M. (2011). Variability in photos of the same face. *Cognition*, *121*(3), 313–323.

Kanwisher, N. (2001). Faces and places: of central (and peripheral) interest. *Nature Neuroscience*, *4*(5), 455–456.

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 : The Official Journal of the Society for Neuroscience*, *17*(11), 4302–4311.

Kay, K. N., Weiner, K. S., & Grill-Spector, K. (2015). Attention Reduces Spatial Uncertainty in Human Ventral Temporal Cortex. *Current Biology*, 1–6.

Kayaert, G., Biederman, I., & Vogels, R. (2003). Shape tuning in macaque inferior temporal cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *23*(7), 3016–3027.

Kiani, R., Esteky, H., Mirpour, K., & Tanaka, K. (2007). Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *Journal of Neurophysiology*, *97*(6), 4296–4309.

Kohler, P. J., Clarke, A., Yakovleva, A., Liu, Y., & Norcia, A. M. (2016). Representation of Maximally Regular Textures in Human Visual Cortex. *J Neurosci*, *36*(3), 714–729.

Konkle, T., & Oliva, A. (2012). A Real-World Size Organization of Object Responses in Occipitotemporal Cortex. *Neuron*, *74*(6), 1114–1124.

Kourtzi, Z., & Connor, C. E. (2011). Neural Representations for Object Perception: Structure, Category, and Adaptive Coding. *Annual Review of Neuroscience*, *34*(1), 45–67.

Kourtzi, Z., & Kanwisher, N. (2001). Representation of Perceived Object Shape by the Human Lateral Occipital Cortex. *Science2*, *293*(5534), 1506–1509.

Kourtzi, Z., & Welchman, A. E. (2015). Adaptive shape coding for perceptual decisions in the human brain. *Journal of Vision*, *15*(7), 1–9. https://doi.org/10.1167/15.7.2.doi

Krekelberg, B., Boynton, G. M., & van Wezel, R. J. A. (2006). Adaptation: from single cells to BOLD signals. *Trends in Neurosciences*, *29*(5), 250–256.

Kriegeskorte, N. (2015). Deep neural networks: a new framework for modelling biological vision and brain information processing. *Annual Review of Vision Science*, *1*, 417–446.

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*(November), 4.

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.

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., & Baker, C. I. (2009). Circular analysis in systems neuroscience: The dangers of double dipping. *Nature Neuroscience*, *12*(5), 535–540.

Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, *23*(11), 571–579.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition & Emotion*, *24*(8), 1377–1388.

Larsson, J., Solomon, S. G., & Kohn, A. (2015). fMRI adaptation revisited. *Cortex*, *80*, 154–160.

Lehky, S. R., Kiani, R., Esteky, H., & Tanaka, K. (2014). Dimensionality of object representations in monkey inferotemporal cortex. *Neural Computation*, *1872*(10), 1840–1872.

Lescroart, M. D., Stansbury, D. E., & Gallant, J. L. (2015). Fourier power, subjective distance, and object categories all provide plausible models of BOLD responses in scene-selective visual areas. *Frontiers in Computational Neuroscience*, *9*.

Levitt, J., Kiper, D., & Movshon, J. A. (1994). Receptive Fields Anf Functional Architecture of Macaque V2. *Journal of Neurophysiology*, *71*(6), 2517–2542.

Levy, I., Hasson, U., Avidan, G., Hendler, T., & Malach, R. (2001). Center – periphery organization of human object areas. *Nature Neuroscience*, *4*(5), 533–539.

Li, L., Miller, E. K., & Desimone, R. (1993). The Representation of Stimulus-Familiarity in Anterior Inferior Temporal Cortex. *Journal of Neurophysiology*, *69*(6), 1918–1929.

Liu, H., Agam, Y., Madsen, J., & Kreimen, G. (2009). Timing, timing, timing: Fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron*, *62*(2), 281–290.

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.

Loschky, L. C., & Larson, A. M. (2010). The natural/man-made distinction is made before basic-level distinctions in scene gist processing. *Visual Cognition*, *18*(4), 513–536.

Loschky, L. C., Sethi, A., Simons, D. J., Pydimarri, T. N., Ochs, D., & Corbeille, J. L. (2007). The importance of information localization in scene gist recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(6), 1431–1450.

Lowe, D. G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, *60*(2), 91–110.

MacEvoy, S. P., & Epstein, R. a. (2007). Position selectivity in scene- and object-responsive occipitotemporal regions. *Journal of Neurophysiology*, *98*(4), 2089–2098.

Mahon, B. Z., Anzellotti, S., Schwarzbach, J., Zampini, M., & Caramazza, A. (2009). Category-Specific Organization in the Human Brain Does Not Require Visual Experience. *Neuron*, *63*(3), 397–405.

Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., … Tootell, R. B. H. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Neurobiology*, *92*, 8135–8139.

Marr, D., & Hildreth, E. (1980). Theory of Edge Detection. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *207*(1167), 187–217.

Martin, A., Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (1996). Neural correlates of category-specific knowledge. *Nature*, *379*(3566), 649-652.

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.

McKone, E., Kanwisher, N., & Duchaine, B. C. (2007). Can generic expertise explain special processing for faces? *Trends in Cognitive Sciences*, *11*(1), 8–15.

McNeil, J. E., & Warrington, E. K. (1993). Prosopagnosia: A face-specific disorder. *The Quarterly Journal of Experimental Psychology Section A*, *46*(1), 1–10.

Mendez, M. F., & Cherrier, M. M. (2003). Agnosia for scenes in topographagnosia. *Neuropsychologia*, *41*(January 2002), 1387–1395.

Miller, E. K., Li, L., & Desimone, R. (1991). A Neural Mechanism for Working and

Recognition Memory in Inferior Temporal Cortex. *Science*, *254*, 1377–1379.

Milner, A. D., & Goodale, M. A. (1995). *The Visual Brain in Action*. Oxford: Oxford University Press.

Milner, A. D., & Goodale, M. A. (2008). Two visual systems re-viewed. *Neuropsychologia*, *46*(2008), 774–785.

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.

Mur, M., Bandettini, P. A., & Kriegeskorte, N. (2009). Revealing representational content with pattern-information fMRI - An introductory guide. *Social Cognitive and Affective Neuroscience*, *4*(1), 101–109.

Mur, M., & Kriegeskorte, N. (2014). What's there, distinctly, when and where? *Nature Neuroscience*, *17*(3), 332–333.

Myers, A., & Sowden, P. T. (2008). Your hand or mine? The extrastriate body area. *NeuroImage*, *42*(4), 1669–1677.

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.

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.

Nasr, S., & Tootell, R. B. H. (2012). A Cardinal Orientation Bias in Scene-Selective Visual Cortex. *Journal of Neuroscience*, *32*(43), 14921–14926.

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A Toolbox for Representational Similarity Analysis. *PLoS Computational Biology*, *10*(4), e10003553.

Nosofsky, R. M. (1986). Attention, Similarity, and the Identification-Categorization Relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–57.

O'Craven, K. M., & Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stiimulus-specific brain regions. *Journal of Cognitive Neuroscience*, *12*, 1013–1023.

O'Toole, A. J., Jiang, F., Abdi, H., & Haxby, J. V. (2005). Partially distributed representations of objects and faces in ventral temporal cortex. *Journal of Cognitive Neuroscience*, *17*(4), 580–590.

Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the United States of America*, *87*(24), 9868–9872.

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., Baker, C. I., DiCarlo, J. J., & Kanwisher, N. G. (2006). Discrimination Training Alters Object Representations in Human Extrastriate Cortex. *Journal of Neuroscience*, *26*(50), 13025–13036.

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.

Op de Beeck, H., Wagemans, J., & Vogels, R. (2001). Inferotemporal neurons represent low-dimensional configurations of parameterized shapes. *Nature Neuroscience*, *4*(december), 1244–1252.

Oppenheim, A. V., & Lim, J. S. (1981). Importance of Phase in Signals. *Proceedings of the IEEE*, *69*(5), 529–541.

Parvizi, J., Jacques, C., Foster, B. L., Withoft, N., Rangarajan, V., Weiner, K. S., & Grill-Spector, K. (2012). Electrical Stimulation of Human Fusiform Face-Selective Regions Distorts Face Perception. *Journal of Neuroscience*, *32*(43), 14915–14920.

Pasupathy, A., & Connor, C. E. (2002). Population coding of shape in area V4. *Nature Neuroscience*, *5*(12), 1332–1338.

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.

Peirce, J. W. (2015). Understanding mid-level representations in visual processing vision. *Journal of Vision*, *15*(7), 1–9.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442.

Pitcher, D., Charles, L., Devlin, J. T., Walsh, V., & Duchaine, B. (2009). Triple Dissociation of Faces, Bodies, and Objects in Extrastriate Cortex. *Current Biology*, *19*(4), 319–324.

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.

Ponce, J., Sturmfels, B., & Trager, M. (2017). Congruences and Concurrent Lines in Multi-View Geometry arXiv : 1608 . 05924v2 [math . AG] 25 Dec 2016. *Advances in Applied Mathematics*, *88*, 62–91.

Portilla, J., & Simoncelli, E. P. (2000). Aparametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision*, *40*(1), 49–71.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology Human Learning and Memory*, *2*(5), 509–522.

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.

Psalta, L., Young, A. W., Thompson, P., & Andrews, T. J. (2014). The thatcher illusion reveals orientation dependence in brain regions involved in processing facial expressions. *Psychological Science*, *25*(1), 128–136.

Rajimehr, R., Devaney, K. J., Bilenko, N. Y., Young, J. C., & Tootell, R. B. H. (2011). The “parahippocampal place area” responds preferentially to high spatial frequencies in humans and monkeys. *PLoS Biology*, *9*(4), e1000608.

Rangarajan, V., Hermes, D., Foster, B. L., Weiner, K. S., Jacques, C., Grill-Spector, K., & Parvizi, J. (2014). Electrical Stimulation of the Left and Right Human Fusiform Gyrus Causes Different Effects in Conscious Face Perception. *Journal of Neuroscience*, *34*(38), 12828–12836.

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 : The Official Journal of the Society for Neuroscience*, *34*(26), 8837–8844.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, *2*(11), 1019–1025.

Ritchie, J. B., Bracci, S., & Op de Beeck, H. (2017). Avoiding illusory effects in representational similarity analysis: What (not) to do with the diagonal. *NeuroImage*, *148*(January), 197–200.

Robbins, R., & McKone, E. (2007). No face-like processing for objects-of-expertise in three behavioural tasks. *Cognition*, *103*(1), 34–79.

Rolls, E. T., Baylis, G. C., Hasselmo, M. E., & Nalwa, V. (1989). The effect of learning on the face selective responses of neurons in the cortex in the superior temporal sulcus of the monkey. *Experimental Brain Research*, *76*(1), 153–164.

Rossion, B., Hanseeuw, B., & Dricot, L. (2012). Defining face perception areas in the human brain: A large-scale factorial fMRI face localizer analysis. *Brain and Cognition*, *79*(2), 138–157.

Rubin, N. (2001). Figure and ground in the brain. *Nature Neuroscience*, *4*(9), 857–858.

Sadeh, B., Pitcher, D., Brandman, T., Eisen, A., Thaler, A., & Yovel, G. (2011). Stimulation of category-selective brain areas modulates ERP to their preferred categories. *Current Biology*, *21*(22), 1894–1899.

Schwarzkopf, D. S., & Rees, G. (2011). Pattern classification using functional magnetic resonance imaging. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(5), 568–579.

Schwarzlose, R. F., Swisher, J. D., Dang, S., & Kanwisher, N. (2008). The distribution of category and location information across object-selective regions in human visual cortex. *Proceedings of the National Academy of Sciences*, *105*(11), 4447–4452.

Shinkareva, S. V, Mason, R. a, Malave, V. L., Wang, W., Mitchell, T. M., & Just, M. A. (2008). Using FMRI brain activation to identify cognitive states associated with perception of tools and dwellings. *PloS One*, *3*(1), e1394.

Sigman, M., Cecchi, G. A., Gilbert, C. D., & Magnasco, M. O. (2001). On a common circle : Natural scenes and Gestalt rules, *2000*, 1935–1940.

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. *Journal of Neuroscience*, *35*(34), 11921–11935.

Slotnick, S. D., Thompson, W. L., & Kosslyn, S. M. (2005). Visual mental imagery induces retinotopically organized activation of early visual areas. *Cerebral Cortex*, *15*(10), 1570–1583.

Sobotka, S., & Ringo, J. L. (1993). Brain Research in unit responses from inferotemporal cortex. *Experimental Brain Research*, *96*, 28–38.

Spiridon, M., & Kanwisher, N. (2002). How distributed is visual category information in human occipito-temporal cortex? An fMRI study. *Neuron*, *35*(6), 1157–1165.

Stojanoski, B., & Cusack, R. (2014). Time to wave good-bye to phase scrambling: creating controlled scrambled images using diffeomorphic transformations. *Journal of Vision*, *14*(12), 6-16.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., … Rabinovich, A. (2015). Going deeper with convolutions. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1–9.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience*, *19*, 109–139.

Thomson, M. G. A. (1999). Higher-order structure in natural scenes. *Journal of the Optical Society of America, A*, (July), 1549–1553.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*.

Tsao, D. Y., Freiwald, W. A., Tootell, R. B. H., & Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science*, *311*(9), 670–674.

Ullman, S. (1998). Three-dimensional object recognition based on the combination of views. *Cognition*, *67*(1–2), 21–44.

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.

Van Essen, D. C., Anderson, C. H., & Felleman, D. J. (1992). Information Processing in the Primate Visual System : An Integrated Systems Perspective. *Science*, *255*, 419–424.

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.

Vogels, R. (1999). Categorization of complex visual images by rhesus monkeys . Part 2 : single-cell study. *European Journal of Neuroscience*, *11*, 1239–1255.

Vuilleumier, P., Henson, R. N., Driver, J., & Dolan, R. J. (2002). Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. *Nature Neuroscience*, *5*(5), 491–499.

Wandell, B. A., Dumoulin, S. O., & Brewer, A. A. (2007). Visual field maps in human cortex. *Neuron*, *56*(2), 366–383.

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.

Warrington, E. K. (1975). The selective impairment of semantic memory. *Q. J. Exp. Psychol. 27,* 635-657.

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., Hymers, M., Hartley, T., & Andrews, T. J. (2016). Patterns of neural response in scene-selective regions of the human brain are affected by low-level manipulations of spatial frequency. *NeuroImage*, *124*(2016), 107–117.

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.

Weiner, K. S., Golarai, G., Caspers, J., Chuapoco, M. R., Mohlberg, H., Zilles, K., … Grill-Spector, K. (2014). The mid-fusiform sulcus: A landmark identifying both cytoarchitectonic and functional divisions of human ventral temporal cortex. *NeuroImage*, *84*(2014), 453–465.

Wilson, H. R., & Wilkinson, F. (2015). From orientations to objects : Configural processing in the ventral stream. *Journal of Vision*, *15*(7), 1–10.

Wolbers, T., Klatzky, R. L., Loomis, J. M., Wutte, M. G., & Giudice, N. A. (2011). Modality-independent coding of spatial layout in the human brain. *Current Biology*, *21*(11), 984–989.

Wyatte, D., Curran, T., & O’Reilly, R. (2012). The Limits of Feedforward Vision: Recurrent Processing Promotes Robust Object Recognition when Objects Are Degraded. *Journal of Cognitive Neuroscience*, *24*(11), 2248–2261.

Xu, X., Yue, X., Lescroart, M. D., Biederman, I., & Kim, J. G. (2009). Adaptation in the fusiform face area (FFA): Image or person? *Vision Research*, *49*(23), 2800–2807.

Yue, X., Tjan, B. S., & Biederman, I. (2006). What makes faces special ? *Vision Research*, *46*, 3802–3811.

Ziemba, C. M., Freeman, J., Movshon, J. A., & Simoncelli, E. P. (2016). Selectivity and tolerance for visual texture in macaque V2. *Proceedings of the National Academy of Sciences*, 201510847.