



On the partnership between neural representations of object categories and visual features in the ventral visual pathway



Stefania Bracci¹, J. Brendan Ritchie¹, Hans Op de Beeck*

Brain and Cognition, KU Leuven, Leuven, Belgium

A B S T R A C T

A dominant view in the cognitive neuroscience of object vision is that regions of the ventral visual pathway exhibit some degree of category selectivity. However, recent findings obtained with multivariate pattern analyses (MVPA) suggest that apparent category selectivity in these regions is dependent on more basic visual features of stimuli. In which case a rethinking of the function and organization of the ventral pathway may be in order. We suggest that addressing this issue of functional specificity requires clear coding hypotheses, about object category and visual features, which make contrasting predictions about neuroimaging results in ventral pathway regions. One way to differentiate between categorical and featural coding hypotheses is to test for residual categorical effects: effects of category selectivity that cannot be accounted for by visual features of stimuli. A strong method for testing these effects, we argue, is to make object category and target visual features orthogonal in stimulus design. Recent studies that adopt this approach support a feature-based categorical coding hypothesis according to which regions of the ventral stream do indeed code for object category, but in a format at least partially based on the visual features of stimuli.

1. Introduction

It is manifest from our everyday experience that we categorize the objects we see. For example, if we see a cat walking down the street, we do not see it simply as a mobile shape, with furry texture, but as an animate creature, an animal, a cat, or possibly our own cat. Even in the extreme case of ambiguous stimuli that are intermediate between two categories, we typically still group the stimuli into one or the other category, instead of experiencing some vague, indeterminate perception (Harnad, 1987). Yet, research aimed at localizing this capacity to particular visual brain regions has sometimes been met with skepticism.

Our capacity to visually categorize stands at the crossroads of the traditional divide between perception and cognition and depends on the interface between sensory processing of external retinal inputs and semantic knowledge of object categories (Op de Beeck et al., 2003; Palmeri and Gauthier, 2004). This raises the question of whether categorical representations can be identified within later stages of visual processing, or are solely the domain of higher cognition. The availability of these theoretical alternatives can influence the interpretation of neuroscientific findings. In particular, it is tempting to hypothesize that the apparent specialization of a visual brain region for representing an abstract property like object category is instead explainable by visual

properties of stimuli. For example, animals and tools differ in animacy, but depending on the exact stimuli that one chooses, they may also differ in shape, color, and texture. In which case, differential neural selectivity for the stimuli could reflect a difference in object category, or visual features (Fig. 1). Indeed, the first results showing object category selectivity in the tuning of neurons in the inferior temporal (IT) cortex of monkeys (Gross et al., 1972) were initially met with such skepticism (for historical discussion, see Gross, 2008).

This same skepticism extends to neuroimaging research investigating object category representations in the ventral visual pathway, which is generally considered to be the neural loci of our capacity to categorize, and is more generally involved in the processing of stable features or dimensions of objects available in the information passed from the retina to early visual regions (Kravitz et al., 2013; Mishkin et al., 1983). Initial fMRI studies revealed the existence of several cortical regions in the pathway that respond preferentially to objects of particular categories, such as the fusiform face area (FFA; Kanwisher et al., 1997), extrastriate body area (EBA; Downing et al., 2001), parahippocampal place area (PPA; Epstein and Kanwisher, 1998), and visual word form area (VWFA; Cohen et al., 2000), to name some of the most well-known regions. In recent years advanced experimental designs and multivariate pattern analyses (MVPA)

* Correspondence to: Laboratorium voor Biologische Psychologie, Tiensestraat 102 - bus 3714, 3000 Leuven, Belgium.

E-mail address: hans.opdebeeck@kuleuven.be (H.O. de Beeck).

¹ Joint first authors.

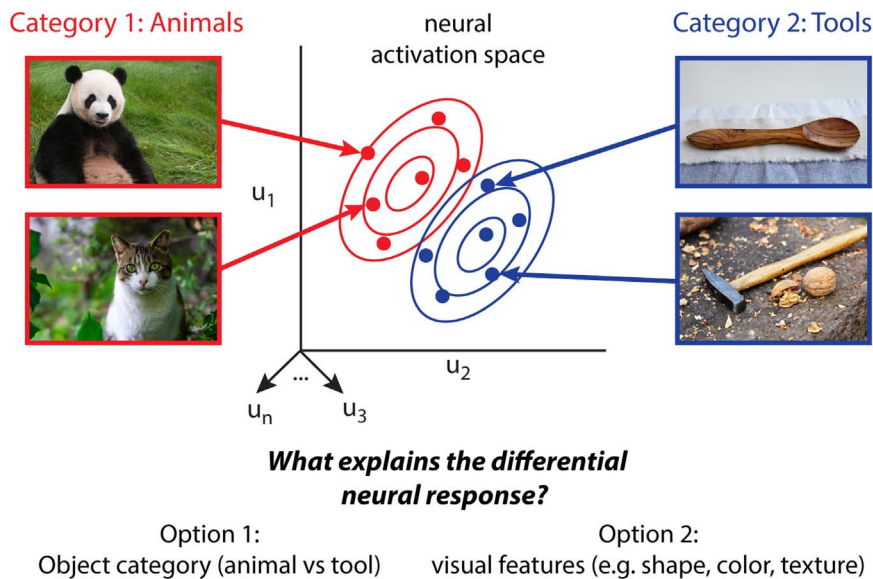


Fig. 1. Competing explanations for apparent categorical effects in the ventral pathway as revealed using MVPA. Stimuli from separate object categories (animals and tools) might produce different patterns of responses in a brain region as reflected in a (hypothetical) N dimensional feature space constructed from neural data from multiple recording units, u_1, \dots, u_n (e.g. electrodes for cellular recordings or voxels from a region of interest for fMRI). When the stimuli also tend to vary in their visual features (e.g. low aspect ratio shape and white-patched furry texture vs. high aspect ratio shape and smooth wood-colored texture), this raises the question of whether the discriminability of the neural patterns for the animal and tool stimuli are explained by the difference in category membership (Option 1) or visual features (Option 2).

techniques, which allow for a more refined exploration of the factors that underlie the neural selectivity, have only intensified the debate. On the one hand, a sizeable body of research has shown that neural activity patterns in regions of this pathway exhibit category selectivity (for review, see Grill-Spector and Weiner, 2014). These results have been interpreted as revealing aspects of neural coding for object categories in visual cortex. On the other hand, recent studies suggest these pattern responses may be heavily feature dependent, raising the question of whether these apparent category effects can in fact be explained by sensitivity to lower-level visual features (Andrews et al., 2015; Baldassi et al., 2013). In which case, coding in these regions may be for visual features of objects, rather than category membership.

In what follows we evaluate different categorical and featural coding hypotheses for regions of the ventral pathway. In Section 2, we briefly review the kinds of MVPA results that have been taken as evidence of either category selectivity or feature dependence. In Section 3, we differentiate between a number of coding hypotheses and their predictions regarding MVPA results. Crucially, we argue that these hypotheses are not necessarily distinguished by whether they predict feature dependence in the pattern responses. Instead, they can be distinguished by whether they predict *residual categorical effects* in these responses after one takes account of the feature dependence. In Section 4, we then turn to a critical discussion of existing MVPA results in light of the available hypotheses, and suggest that evidence for residual categorical effects comes from studies that either: (i) model out the contribution of different visual features to categorical effects; or (ii) also use stimulus sets in which object category and target visual features are orthogonal. While the former is more common, we emphasize the importance of the latter approach. We further argue that the results of MVPA studies that make object category and target visual features orthogonal support the presence of categorical coding in regions of the ventral pathway. In Section 5, we consider what other lines of converging neuroscientific evidence are relevant to deciding between coding hypotheses. In Section 6, we end with some morals and suggestions for future directions.

2. Using MVPA to investigate object category representations in the ventral visual pathway

We begin with some clarificatory remarks about the nature of object categorization and the theoretical rationale for using MVPA to investigate its neural basis.

By visual “object categorization” we mean the process of matching a

representation of an object constructed online to a categorical representation from long-term memory, such that the categorical property that is represented is attributed to the object perceived in the environment. We take categories to reflect natural “kinds of things” in the environment, which may be taxonomic (e.g. biological or artefactual kinds), or functional (e.g. ways in which an object can be manually manipulated, or whether it presents a possible food source or threat). Many of the specific categories we mention, such as faces, bodies, and tools, have clear ecological significance to how we make sense of, navigate, and interact with the objects in the world around us. Category membership, in turn, is presumed to be determined by a property, or set of properties, instantiated by member objects—though intuitively we might only have a dim idea of what these underlying properties might be. What is crucial is that the shared properties are abstract in comparison to more basic visual properties of the stimuli, such as color, texture, or even shape, even though this distinction is not always very clear-cut because visual features also vary in complexity.

The space of object categories is highly complex, with multiple hierarchical levels and a multidimensional similarity structure per level, as can be seen when representing a familiar object like a cat: it can be represented as an animate object, a mammalian animal, or a feline, with each level admitting of different dimensions of variation. Even though univariate neuroimaging designs that contrast individual categories have been able to reveal fundamental properties of the neural system underlying object categorization, this initial approach is probably insufficient to fully characterize the neural representations that sustain object categorization. In contrast, MVPA has been shown to be well suited to characterize these neural representations in their full complexity, including the study of multiple hierarchical levels and multidimensional structures (e.g. Edelman et al., 1998; Hanson et al., 2004; Kiani et al., 2007; Kriegeskorte et al., 2008a).

The analysis of distributed neural patterns with MVPA also seems to be an appropriate method given the prevalent working hypothesis in cognitive neuroscience that the brain employs population coding, where stimulus information is encoded in distributed patterns of neural activity across a brain region (Pouget et al., 2000; Panzeri et al., 2015). Thought of geometrically, the hypothesis entails that each distributed response pattern can be characterized as a point in a high dimensional “activation space”, where dimensions reflect different neuronal units (e.g. individual neurons or cortical columns). MVPA is suitable for investigating this sort of neural coding if we assume both that neuroimaging techniques coarsely measure these encoding neural patterns in a region, and that the activation spaces constructed from these patterns

have some fidelity to the underlying representational “geometry” (Haxby et al., 2014; Kriegeskorte and Kievit, 2013; Op de Beeck, Haushofer, and Kanwisher, 2008). However, we should note that the use of MVPA does not require a blind faith that these assumptions are correct (Davis et al., 2014; Spiridon and Kanwisher, 2002).

So described, MVPA inherits the dependence of human fMRI in general upon the existence of neural maps in which neurons with similar functional properties are spatially clustered together in cortex. Such clustering is necessary to generate patterns of activity that are differentiable across stimulus conditions, and can be detected both with cellular recordings and at the coarse spatial scale of fMRI. Crucially, the detectability of such maps with fMRI should not be taken for granted, since a brain region may code for properties of stimuli in a manner that is not amenable to MVPA (Dubois et al., 2015), or that produces biases or distortions in measurement depending on the spatial scale of the encoding patterns (Kriegeskorte and Diedrichsen, 2016; Op de Beeck, 2010). But at least when it comes to category-related selectivity, there is some evidence that the results of fMRI do coarsely measure spatial patterns detected with cellular recordings for object stimuli (Issa et al., 2013).

In the rest of this section we review a selection of the types of MVPA results that have provided evidence of either category selectivity or feature dependence in the ventral visual pathway, which we take to include a complex array of inter-connected brain regions across lateral occipitotemporal cortex (LOTc), IT cortex, and medial temporal cortex (Kravitz et al., 2013). This includes posterior portions of the fusiform, and parahippocampal gyri in humans—also known as ventral temporal cortex (VTC)—that is the focus of a substantial portion of human fMRI research on object categorization (Grill-Spector and Weiner, 2014).

2.1. Evidence for categorical representations

There are two kinds of MVPA results that have been interpreted as reflecting object category representation in regions of the ventral pathway.

First, if the brain uses a population code to represent object categories in regions of the ventral pathway, then one prediction is that neural patterns for stimuli of the same category should cluster together in the activation spaces of these regions, and hence be reliably discriminable from each other. For example, since we visually represent *animal* and *tool* as distinct categories, in theory the neural response patterns for exemplars of these two categories should be differentiable in some region of visual cortex. MVPA provides one prominent means of determining whether neural patterns can be so differentiated, thus revealing evidence of *categorical information* in the pattern responses of a brain region (Kriegeskorte and Bandettini, 2007).

Many early studies using MVPA showed that information for a wide variety of object categories can be decoded from pattern activity across regions of the ventral stream. In a seminal fMRI study, Haxby et al. (2001) showed that neural patterns in VTC objects of familiar categories could reliably be differentiated. Similarly, the first studies using classifiers with human fMRI showed they could be trained to discriminate neural patterns in VTC for faces, houses and chairs (Carlson et al., 2003), and a wide variety of living/non-living, common/uncommon, and large/small objects (Cox and Savoy, 2003). Hung et al. (2005) applied the same machine learning methods to recordings from anterior IT cortex in monkeys showing that stimulus patterns could be reliably grouped into multiple categories achieving peak classifier performance at just 125 ms post-stimulus onset. Similarly, Liu et al. (2009) applied classifiers to intracranial recordings for similar object category stimuli from sites across the human brain, and found that object category could be decoded as soon as 100 ms post-stimulus onset from VTC.

MVPA has also been used to investigate the information that univariate-defined regions of the ventral pathway carry about their region-defining category as well as other object categories. Early studies

suggested that faces and places were more decodable in FFA and PPA respectively, in comparison to other categories (O’Toole et al., 2005; Spiridon and Kanwisher, 2002; though see Op de Beeck et al., 2010), with similar results in other areas such as adjacent hand and body regions of LOTc (Bracci et al., 2012). The same pattern of results has also been observed in both face and body selective regions of medial temporal cortex in humans and monkeys using fMRI (Pinsk et al., 2009). Finally, MVPA has also been used to discriminate between real and pseudo-words in VWFA (Baeck et al., 2015).

The second type of finding is premised on the idea that the relationships between neural patterns for objects in ventral pathway regions reflect the organizational relations between categories. For example, while at one level cats and dogs are different kind of objects, they are both instances of domestic pets, mammals, and animals more generally. If a region codes for these inter-category relations, then a *categorical organization* might also be extractable from how neural patterns for different exemplars cluster in activation space (Haxby et al., 2014; Jordan et al., 2015; Kriegeskorte and Kievit, 2013).

Several studies have provided evidence of different forms of categorical organization in regions of the ventral pathway. One prominent finding is that patterns of neural activity in human and monkey IT cluster based on the superordinate animate-animate distinction, and that for animate exemplars the neural patterns further divide into subordinate categories (Kiani et al., 2007; Kriegeskorte et al., 2008b; cf. Hanson et al., 2004). Furthermore, this organization in human IT is similar to that found in monkey IT (Kriegeskorte et al., 2008b), and may reflect a continuum of animacy rather than a dichotomous representation in the ventral pathway (Sha et al., 2015). The combination of fMRI with MEG has also allowed temporally resolved localization to the ventral pathway for many of these same divisions of animacy and subordinate (e.g. faces and bodies) categories (Cichy et al., 2014, 2016).

The same methods have also been used to reveal other forms of categorical organization. Connolly et al. (2012) found that the clustering of neural patterns in VTC and LOTc for images of insects, birds, and primates reflected a natural taxonomic hierarchy between categories. Going further, Connolly et al. (2016) also found a representation of the dimensions of animacy versus dangerousness in numerous ventral pathway regions. Other categorical divisions are not taxonomic in nature, but have also been revealed with MVPA. The same analyses have revealed an organization reflecting a function-related relationship for body parts and tools in LOTc and VTC (Bracci et al., 2015; Bracci and Peelen, 2013). They have also been used to investigate the categorical organization in more circumscribed brain regions, including those defined by univariate methods, such as FFA, PPA, and EBA (e.g. Op de Baeck et al., 2010). Furthermore, many of the studies mentioned above, which report effects of categorical information in regions of the ventral pathway, have also investigated forms of categorical organization.

2.2. The issue of feature dependence

The above MVPA results have been taken to provide converging evidence that ventral pathway regions implement representations of object category. However, a reasonable concern, as with any study using visual stimuli, is that some of these effects might reflect not the abstract property of category membership, but more low-level visual differences between stimuli that are highly correlated with category, and are not controlled for (Cox and Savoy, 2003). For example, if a region shows discriminable pattern responses for animal and tool stimuli, shape might drive the response if there is a consistent difference in aspect ratio between the categorical stimuli (Fig. 1). It is for this reason that many studies that report categorical effects contrast results from ventral pathway regions with those from early visual cortex (EVC)—typically primary visual cortex (V1)—or test whether the categorical effect they report can be account for by low-level image

properties.

For example, many studies showing evidence of categorical organization for animacy have used color stimuli, which presents a possible confound (Kiani et al., 2007; Kriegeskorte et al., 2008b; Sha et al., 2015). Of course, color is also a highly diagnostic for differentiating between members of different categories (compare fruits and tools)—a point we return to below. However, plausibly if a region contains representations of different categories neural responses should presumably not be accounted for solely by stimulus color (e.g. one can still tell the difference between an animal and a tool even under lighting conditions in which color is difficult to determine). To address this sort of concern, Kriegeskorte et al. (2008b) compared a color model (along with several other models) to neural responses from human and monkey IT. They found that the model did not capture the relationships between pattern responses. Similarly, Haxby et al. (2001, Appendix) reanalyzed the data of Ishai et al. (1999) and found that VTC could discriminate neural patterns both for natural images, and line drawings of exemplars, suggesting representations not intrinsically tied to object texture. While color is often easy to control for (by presenting stimuli in grey-scale) and perhaps texture as well, other image properties and visual features have also been identified as potentially driving neural responses in regions of the ventral visual pathway.

Shape provides one natural confound, since objects of the same category often have the same shape. When using computer-generated novel objects, activity patterns in LOTC have been shown to correlate with a model of the perceived shape of these objects (Op de Beeck et al., 2008b). Similar findings have been found in monkeys using single-unit recordings and fMRI (Op de Beeck et al., 2001; Yamane et al., 2008). In one study, Baldassi et al. (2013) sought to directly compare the contributions of object category and shape to the response patterns in monkey IT. They presented a large number of greyscale object stimuli to monkeys that included 94 category pairs (e.g. two butterflies, two fish hooks) while recording from sites in monkey IT. They found that shape properties of stimuli, rather than category, better accounted for the relations between neural patterns based on clustering analysis. Based on these findings, Baldassi et al. concluded that IT neurons primarily represent visual information related to shape, rather than object categories. These results, and their interpretation, are seemingly in contrast with those of Kiani et al. (2007) and Kriegeskorte et al. (2008b) described above. (However, it is also worth considering that one issue with many studies investigating object categorization in primates is that the animals have little to no understanding of the images they see. From this vantage point, the results of Baldassi et al. may not be wholly surprising).

Low-level image properties provide another possible confound. As mentioned above, many of the studies using MVPA that report category selectivity take some account of this possibility. However other results have been interpreted as showing that seemingly categorical effects are actually driven by low-level properties. Rice et al. (2014) constructed a model for images of several familiar categories using the GIST model of scene statistics (Oliva and Torralba, 2001), which was compared with the neural patterns for multiple anatomical regions in the ventral pathway measured with human fMRI. Their results showed that the relations between the neural patterns for different categories could be predicted by the GIST model (but see Wardle and Ritchie, 2014). Similarly, Watson et al. (2016b) found that spectral and spatial GIST models were able to predict neural responses in VTC. Also, Coggan et al. (2016) found that the neural patterns for scrambled versions of object stimuli in VTC were positively correlated with the neural responses for the unscrambled images, despite the fact that the scrambled images were unrecognizable. The authors interpreted this result as providing evidence that low-level properties of images differentiate between category-related neural patterns in VTC even in the absence of any overt representation of object category. Based on such findings, Andrews et al. (2015) propose that pattern responses in the ventral pathway are better explained by image low-level properties than object category.

These studies share an underlying rationale: if the clustering of neural patterns for object exemplars is accounted for by some of these stimulus properties—that is, if the responses are heavily *feature dependent*—then this could challenge the hypothesis that the relevant region codes for object category. Furthermore, these results raise the question of whether apparent categorical effects are in fact featural ones.

3. Contrastive coding hypotheses for the ventral visual pathway

The findings we have reviewed point to conflicting perspectives about the functional organization of the ventral visual pathway. How are these findings to be evaluated? To answer this question, we believe one must first distinguish between the alternative hypotheses that are available, and articulate their predictions with respect to the effects that can be revealed using MVPA.

At issue is the nature of the neural population coding in regions of the ventral pathway. In effect, there are two broad alternative hypotheses under consideration, which differ most fundamentally in their content (as opposed to format), or the information they make explicit (Grill-Spector and Weiner, 2014; Marr, 1982): object categories or visual features. In general terms we will call these *categorical* and *featural* coding hypotheses, respectively. Different versions of these coding hypotheses provide a means of conceptualizing the dialectic from the previous section, and how we should interpret categorical effects and evidence of feature dependence (Fig. 2). For any given coding hypothesis we can ask which of these MVPA results it does (and does not) predict.

3.1. Does the coding hypothesis predict category selectivity?

Consider first what sort of featural coding hypothesis is ruled out when there is evidence of category selectivity in a region like VTC, as there is for discriminating animals vs tools. As described above, many studies compare the neural pattern response in this region to models of low-level visual cortex, or compare categorical models to early visual cortex. When this is done, studies tend to find a difference in decoding ability between EVC and VTC for low-level features relative to more high-level and categorical information (Connolly et al., 2012; Kriegeskorte, 2008b; Op de Beeck et al., 2008b; Peelen and Caramazza, 2012). Clearly these results rule out what we may call a *low-level featural* code. Under such a hypothesis, a region contains feature maps for visual properties such as orientation, spatial frequency, or 2D shape, in analogy to the organization of early visual areas.

When applied to regions of the ventral pathway like VTC, a low-level featural code predicts that there should be no categorical effects.

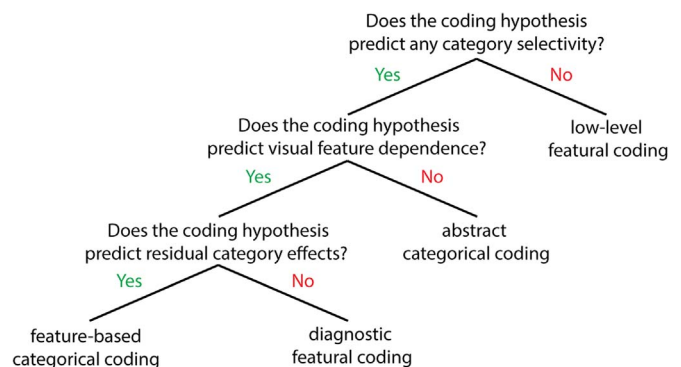


Fig. 2. Decision tree for coding hypotheses. Each branch in the decision tree relates to MVPA results that are (or are not) predicted by the different categorical and featural coding hypotheses. Ultimately, the two most plausible hypotheses, feature-based categorical coding and diagnostic featural coding, are distinguished by whether they predict residual categorical effects after one takes account of the feature-dependence of neural responses in a brain region.

In early visual areas, such as V1, one typically finds poor decoding for stimuli based on object category (Connolly et al., 2012; Kriegeskorte et al., 2008b), but excellent decoding for visual features such as orientation (Haynes and Rees, 2005; Kamitani and Tong, 2005). If an area of the ventral pathway also implements a low-level featural code, then one would predict a similar pattern of results—to put the point simply, MVPA results in later regions of the ventral pathway should be similar to those found in V1. The fact that they are not suggests that a low-level featural code is very likely false when considering regions of the ventral pathway.

While ruling out such an alternative is important, it is doubtful that studies showing feature dependence of pattern responses in regions like VTC are intended to support a low-level featural code, since all will agree that the function and organization of VTC differs from EVC. This raises the question of what a plausible featural code might look like—a question we return to shortly.

3.2. Does the coding hypothesis predict visual feature dependence?

Next, consider what kind of categorical coding hypothesis may be targeted by studies showing evidence of feature dependence, as might be the case if neural patterns for animal and tool stimuli can be discriminated from each other in a brain region, but shape, color and texture are highly correlated with object category in the stimulus set (Fig. 1).

Under what we will call an *abstract* categorical code, the format of the representation is largely independent of visual features of stimuli, in a manner similar to domain-specific hypotheses about concepts (Caramazza and Shelton, 1998; Mahon and Caramazza, 2008). According to these theories, although sensory representations may have a causal influence on the organization of our concepts, they do not determine their format (Mahon, 2015). In the present case, the abstract code may be specific to the visual system, and be a natural consequence of the idea that along stages of the ventral pathway representations become more invariant, or tolerant, across transformations of the viewpoint of a stimulus; or can be recruited by purely symbolic reference to the object (e.g. novel character strings that observers learn to associate with animals and tools). Alternatively, it may be implemented in a region of temporal cortex that receives inputs from multiple sense modalities. In either case, at the extreme, visual features might be entirely dissociable from the categorical property that is being attributed to a stimulus. In which case, the hypothesis does not predict feature dependence in a brain region.

In some cases researchers focusing on the ventral pathway do suggest something like an abstract categorical code, perhaps because a region partially implements our conceptual knowledge. For example, Peelen and Caramazza (2012) found that information about the type of action required by a tool (squeeze or rotate), and where the tool is located (kitchen or garage), was higher in the anterior temporal lobes. In contrast, perceptual and pixelwise information was much lower in these regions. Peelen and Caramazza interpreted their findings as showing that categorical, or even conceptual, representations for tools are implemented in these regions. In another study, Malone et al. (2016) had subjects learn to associate lists of pseudowords with different object categories, and then trained a classifier to discriminate object category based on neural responses to the letter strings. They found that category information, induced by the pseudoword stimuli, could indeed be decoded from the lateral anterior temporal lobe, which they suggest must implement an abstract representation of category.

Evidence of the feature dependence of neural responses would seem to speak against a rigidly abstract view of how particular regions of the ventral pathway encode for object categories. Thus, evidence of feature dependence in VTC runs counter to the predictions of an abstract categorical code, though it does not speak against an abstract code being implemented in more anterior regions of the temporal lobe, as suggested by the results of Peelen and Caramazza (2012) and Malone et al.

(2016). Feature dependence also does not necessarily speak against all categorical coding hypotheses. After all, there must be some relationship between the visual representation of object category and visual features, since any region encoding information about object category will be causally downstream from ones representing more low-level properties. Furthermore, since objects of the same category might often be visually similar (e.g. faces), we might predict a good deal of feature dependence, even in categorical representations. We now turn to a discussion of this issue.

3.3. Does the coding hypothesis predict residual categorical effects?

The coding hypotheses we have considered so far do not necessarily speak directly to the tension between the results described in the previous section. While low-level featural coding might be ruled out by category selectivity in ventral pathway regions, it is something of a null hypothesis in the first place. Although abstract featural coding may be more plausible, it seems unviable for a region like VTC. Fortunately, more plausible featural and categorical coding hypotheses are available.

A more plausible featural coding hypothesis is that a regions represents bundles of visual features that are diagnostic of particular object categories (Andrews et al., 2015; Jozwik et al., 2016; Proklova et al., 2016). Call this as the *diagnostic* featural coding hypothesis. Such a proposal is put forth by Andrews et al. (2015), who suggest that apparently discrete category selective regions in VTC may in fact implement multiple featural maps. Since combinations of particular visual features are indicative of membership in particular object categories, these maps would produce spatially selective, and hence seemingly category specific, patterns of neural response, but would not reflect category-determined functional subdivisions of the ventral pathway.

Importantly, a diagnostic featural code is potentially consistent with apparent categorical effects in a brain region, but holds that it is diagnostic features of stimuli that are driving the differentiation between neural responses. For example shape, which reflects multiple properties, might be represented as distinct, because it is diagnostic of category membership (Baldassi et al., 2013). Thus, rather than implementing categorical representations of categories like animal and tool, perhaps (sub-regions of) VTC contain feature maps of shape, as well as other visual features. It is also plausible that some regions of the ventral pathway show a preferential response to some diagnostic features of object stimuli, as we describe in Section 5.

An alternative hypothesis is a *feature-based* categorical code: VTC and its subregions encode object category and use feature-based representations to do so. For example, early results were interpreted as supporting a distributed code that consisted of featural maps for representing object category (Haxby et al., 2001; O'Toole et al., 2005). There are a number of reasons for why a feature-based categorical code should be considered the default hypothesis when it comes to the representation of categories in the ventral stream.

First, virtually all theories of visual object recognition hold that coding for object categories, while viewpoint-dependent, also achieving a measure of viewpoint invariance, or “tolerance”, with respect to transformations of orientation, illumination, distance, and position in the visual field (Biederman, 2000; Hayward, 2003; Peissig and Tarr, 2007). One can for example readily recognize that an object is an animal (e.g. cat) or tool (e.g. a hammer), even though viewing angle, lighting, and viewing distance might vary considerably. However, any representation of a viewpoint of an object will plausibly be heavily feature dependent, due to the features of an object coded in the viewpoint representation. Therefore, it stands to reason that these theories predict a good deal of feature dependence in the neural representations that underlie our capacity to recognize objects. For example, both familiar and newly learned objects may have “canonical” viewpoints—those that come to mind most easily when we imagine an object and produce lower choice and reaction time errors for

categorization (Blanz et al., 1999). If viewpoint-dependent representations of categories are structured around canonical viewpoints, then the visual features encoded in these viewpoints would make the representations feature-based (Cutzu and Edelman, 1994). These viewpoint-dependent theories of recognition provide theoretical background to research on the neural basis of object categorization (DiCarlo et al., 2012), however a feature-based categorical code need not be unimodal and could also be partially constructed based on inputs from other modalities such as touch or audition—a possibility we return to in Section 5.

Second, there are a number of ways featural information might be carried over into downstream brain regions that code for object category. First, following DiCarlo and Cox (2007) we can think of the population coding of the brain as reflecting manifolds in high-dimensional activation spaces for different regions. Early in the visual system, manifolds for object categories are “tangled”, and the visual system works to untangle them across multiple stages of processing. From this perspective, all the categorical information that needs extracting is tacit in the visual inputs, and the function of the visual system is to make this information explicit (Cox, 2014). But if there is categorical information tacit in the tangled manifolds of early visual regions, then it is not unreasonable to expect that there will be featural information still present when these manifolds are later untangled even though it is not necessarily explicitly coded for in these regions. Alternatively, other more high-level features (e.g. shape) might be computed in parallel with categorical information to help disentangle categorical representations. In which case, some featural information might actually improve in decodability along later stages of visual processing, as has indeed been demonstrated by Hong et al. (2016).

Whatever the details of a feature-based categorical code, it is clear that, like a diagnostic featural code, it is consistent with both the presence of categorical effects, and some level of feature dependence, in ventral pathway regions. The main difference between these hypotheses is that a diagnostic featural code predicts feature dependence based on representational content (it is a code for co-occurring diagnostic features of objects), while feature-based categorical coding predicts feature dependence based on format (it is a code for object category, constructed from representations that also code for visual features). How then do we use MVPA to distinguish between these hypotheses?

The rationale behind testing for feature dependence is that apparent category selectivity might involve a failure to control for the confounding influence of visual features. However, if one does control for particular (supposedly) diagnostic features, and still finds a *residual* categorical effect, then this would seem to provide evidence of a categorical code being implemented. For example, if one introduced greater variation in the toy stimulus set in Fig. 1, so that shape was not such a clear confound, and found that one could still discriminate the neural patterns of the animal and tool stimuli, this would be a residual categorical effect with respect to shape. This provides one way of differentiating between feature-based categorical coding and diagnostic featural coding hypotheses in a brain region using MVPA. For under the latter hypothesis, any categorical effects should ultimately be explainable by a summation of the diagnostic visual features of stimuli.

3.4. Summary

We have considered a few coding hypothesis: low-level featural, abstract categorical, diagnostic featural, and feature-based categorical codes. The relationships between these hypotheses, and their differential predictions regarding the sort of MVPA results we have reviewed, can be presented in terms of a decision tree (Fig. 2). As suggested by the structure of the tree, we believe that the two most plausible alternatives are diagnostic featural coding and feature-based categorical coding, both of which are broadly consistent with effects of category selectivity and feature dependence, but are distinguished by whether they predict

residual categorical effects. Still, our presentation only emphasizes certain relationships between these hypotheses. For example, we do not consider whether two kinds of coding might be presenting in a single brain region. Still, we believe that the contrasts between the hypotheses that we have highlighted are instructive. We now turn to evaluating what kinds of neuroimaging studies test for residual categorical effects.

4. Testing for residual categorical effects

There are two main strategies we focus on for revealing residual categorical effects with MVPA. The most common approach is to compare visual feature models to the pattern responses in a brain region. However, we believe a stronger test is to also explicitly control for featural or categorical effects in a brain region. Here we discuss research that adopts both kinds of strategies.

4.1. Featural models and residual categorical effects

One way to test for residual categorical effects, or feature dependence, is by applying a categorical model and various visual models to neural data, and evaluating the variance accounted for by the different models (Kriegeskorte et al., 2008a; Jordan et al., 2015). If some of the variance in the data accounted for by a categorical model is not accounted for by a featural one, then this suggests that there is a residual categorical effect.

A first consideration is which (of the numerous) featural models should be included in an analysis. If the aim is simply to use the models as a substitute for controlling for particular visual features of the stimuli, then this might warrant comparing a large number of available models, as is done in studies that compare models of low-level vision to neural responses in ventral regions. However, if we want to differentiate between diagnostic featural coding and feature-based categorical coding in a region then the models must be theoretically well motivated. This is not always clearly the case.

For example, the GIST model, which applies Gabor filters of different orientations and spatial frequencies to measure the image statistics at different locations in an image, is appropriate for investigating feature dependence in scene-selective regions, since it is based on a theory of how image statistics are diagnostic for scene categorization (Watson et al., 2014, 2016a). However, it is unclear how one should interpret a model of scene statistics when compared to neural patterns from object-selective regions (e.g. Rice et al., 2014; Watson et al., 2016b). Part of the motivation for the model is to treat each scene image as, in effect, objects of the same (square) shape. However, in studies on object categorization the stimuli are typically natural images of objects presented on a grey background, which is importantly different than the kind of stimuli the model was intended for. And with actual shape boundaries, it is likely that the model largely provides another measure of shape—a point we return to below. Second, as we have pointed out, feature dependence is consistent with feature-based categorical coding, thus revealing a relationship between the GIST model and neural responses in regions like VTC does not clearly speak against this hypothesis.

A similar sort of criticism can be leveled against studies that use semantic feature labeling of images, which is common in psychological research on conceptual understanding. Jozwik et al. (2016) constructed high-dimensional categorical and featural models based on free labeling of the stimuli from Kriegeskorte et al. (2008b). Both models were then correlated with the pattern response for the stimuli in IT. Ultimately, they found that both models explained the same variance in the neural data. One could interpret such a result as showing a failure to find a residual categorical effect in IT, in so far as the categorical model did not significantly explain more variance than the featural model. In their study, the categorical model included properties shared by many stimuli, (e.g. “living/organism”) and names for individual stimuli (e.g. “wolf/fox”). The feature-based model included object parts (“horns”,

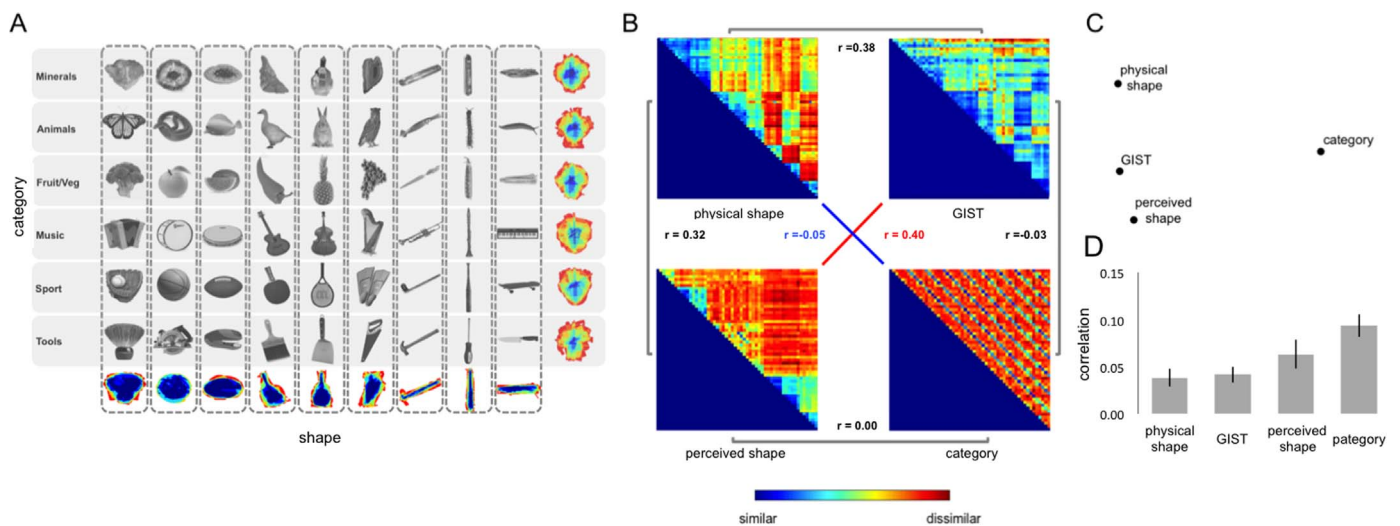


Fig. 3. Analysis of stimuli from Bracci and Op de Beeck (2016). Panel (a) depicts the stimuli from Bracci and Op de Beeck (2016), which involved objects from 6 different categories and 9 different shapes so that category and shape were rendered orthogonal dimensions of the stimulus set. Panel (b) depicts the correlations between a pixel similarity matrix, shape and category judgment similarity matrices, and the GIST model of scene statistics. Panel (c) depicts the 2-dimensional multi-dimensional scaling solution for the similarity between the four matrices in panel (b). As can be seen, the category model does not cluster with the other three. Panel (d) shows the mean correlation between the four matrices and one region of interest from Bracci and Op de Beeck (2016), body and face-selective ventral occipitotemporal cortex (VOTC-body/face).

“neck”), shapes (“cylindrical”), color (“white”, “yellow”), and textures (“wooly”). However, other dimension of the featural model (“dress”, “tree”, “head”, “torso”) could be considered categorical. Similarly, Clarke and Tyler (2014) compared a model of normed “semantic features” (Taylor et al., 2012) to neural response from regions of the ventral pathway. The labeled features of the model included colors (“is green”) and shape (“is round”), but also body parts (“has 4 legs”), and functional properties (“is edible”, “used for music”), and so seem to blur together both featural and categories properties. These semantic models are not constrained by hypotheses about the representation of visual features, but rather by what properties participants can verbally attribute to the stimuli. Therefore, they are also ill-suited to the question of whether ventral pathway regions code for object category vs. diagnostic visual features.

The forgoing speaks to the importance of focusing on features that might be considered diagnostic of particular object categories. Ideally, studies should follow the same rationale as multiple regression, and determine how many different model predictors account for the dependent neural variable. One can also take a neural network modeling approach. In particular, recent studies have incorporated deep neural networks (DNNs) to investigate biological vision (Kriegeskorte, 2015). Several studies have found that the higher layers of networks that have representational spaces similar to those of human and monkey IT perform better at object categorization tasks of natural images, are also better at explaining the activation space from these regions, and seem to be organized based on object categories (Cadieu et al., 2014; Khaligh-Razavi and Kriegeskorte, 2014; Yamins et al., 2014). In contrast lower layers, which are tuned to more visual properties of stimuli, tend to poorly capture the relationships between the neural responses in IT (though it can be difficult to discern precisely what visual features of images the lower layers are tuned to). In so far as it may be difficult to explain the unit responses in the top levels of these networks simply in terms of diagnostic features, and they capture the pattern responses in IT, their organization may point towards the presence of categorical coding in IT.

4.2. Controlling for visual features reveals residual categorical effects

A powerful way to tease apart the contributions of object category and target visual features to a neural response is to make them orthogonal properties of a stimulus set. Any residual categorical effect in the

neural response, as revealed by the modeling approach, is presumably driven by some independence between the categorical and featural information extracted from the stimuli. Making target visual features orthogonal to category membership ensures this independence. Furthermore, if there is no residual categorical effect, then we have better evidence that the visual feature (e.g. shape) that has been made orthogonal was in fact driving the apparent categorical response. This approach has led to much success in research looking at neuronal tuning, where a limited number of visual features are controlled for, and rendered orthogonal to object category (Gross et al., 1969; Tanaka et al., 1991; Yamane et al., 2008; Brincat et al., 2004). Note that a stimulus-based approach is not a substitute for modeling out the contribution of category membership and visual features to a neural response. Rather, where possible the best practice is to use both approaches in conjunction.

Of course, categorical information is typically correlated with many visual features, and in that respect the highly controlled stimulus sets are not representative of our everyday visual experience. One may therefore question the generalizability of such stimulus designs to more ecological settings (Talebi and Baker, 2012). Still, this stimulus-design approach does provide a strong test for comparing feature-based categorical coding and diagnostic featural coding. For if categorical information or organization are present in a brain region, even when target visual properties are made orthogonal to object category, then this would constitute a clear residual categorical effect, and rules out a diagnostic featural coding hypothesis based on these properties.

Recent studies have adopted this stimulus-design approach to differentiating between coding hypotheses by making 2D shape orthogonal from object category. Bracci and Op de Beeck (2016) selected stimuli as part of six separate object categories and nine different shape profiles, such that there was a stimulus for each combination of category and shape (Fig. 3A). This ensured that models of pixel similarity and shape were entirely uncorrelated with a model for object category. These models were correlated with activity patterns from multiple regions of interest. Crucially, in regions of LOTC and VTC both the visual and categorical models were significantly correlated with the neural activity patterns. But since these models were independent, the correlation with the categorical model constituted a residual categorical effect. Other recent studies have adopted the same logic. Proklova et al. (2016) selected animate and inanimate objects that were grouped into three shape clusters, and tested models of both shape and texture. They

found that in regions of the ventral pathway the neural patterns for these shape-matched stimuli could not be accounted for by the visual feature models. In another study combining fMRI and MEG, Kaiser et al. (2016) presented body part (bare hands and clothed torsos) and clothing stimuli (gloves and shirts) that were matched for shape. Using a cross-decoding approach (training on one shape pair, and testing on the other), they found that object category (body vs. clothing) could be decoded bilaterally in LOTC, and at 130–200 ms post-stimulus onset.

These findings have important consequences for the interpretation of results where visual features are highly correlated with object category (Baldassi et al., 2013; Rice et al., 2014). To illustrate this point, we applied the GIST model to the stimuli of Bracci and Op de Beeck (2016). As shown in Fig. 3B–C, using a well-controlled stimulus set, the visual feature model is highly independent from the category model, thus image statistics cannot explain the residual categorical effects observed in the regions of interest (Fig. 3D). Conversely, the GIST model partially correlates with other image visual properties: object perceived shape and object physical (pixel) shape. Together, these results show that residual categorical effects could go largely unnoticed when object category and certain visual features are largely correlated (for an investigation into how these stimuli might be revealing of shape perception, see Kubilius et al., 2016).

The studies we have described have investigated categorical effects in the absence of featural ones, but it is important to recognize that the reverse can also be investigated: featural effects in ventral pathway regions in the absence of categorical ones. A recent study by Coggan et al. (2016) provides an elegant illustration of this idea. fMRI subjects were presented with exemplars of object categories under three conditions: globally and locally scrambled, and unscrambled. In the two scrambled forms, the images were unrecognizable by subjects. By presenting the stimuli in separate runs, increasing in clarity, Coggan et al. (2016) ensured that no categorical representations were recruited by subjects until the unscrambled condition. They then compared neural activity patterns for the stimuli from these conditions in VTC. They found a positive correlations between the neural patterns for the scrambled conditions and the unscrambled condition, suggesting that low-level image properties produce featural effects in VTC even when no categorical representation is recruited.

The design of Coggan et al.'s study is explicitly intended to test against categorical coding in VTC. However a feature-based categorical coding hypothesis is in fact consistent with their results. First, in their stimulus set object category and visual properties were not dissociated, thus correlation between neural patterns for scrambled and unscrambled images could still reflect properties common to images within the same category (e.g., round shape of faces). Second, as they acknowledge, the explainable variance in the neural patterns for intact images could not be entirely explained by the correspondence with the scrambled images. Thus these results also suggest the possible presence of residual categorical effects.

Finally, making category orthogonal to visual features is also relevant to studies using DNNs. Some studies used largely uncontrolled stimuli, like Khaligh-Razavi and Kriegeskorte (2014) who analyzed the data of Kiani et al. (2007) and Kriegeskorte et al. (2008b). However other studies rely on a stimulus design motivated by the viewpoint invariance of object recognition. These stimuli of digitally rendered 3D objects are presented at different locations, scales, and orientations, against random scene backgrounds are designed to make category orthogonal to many incidental visual features created by consistency of viewpoint, and provide a strong test of DNNs on object categorization tasks (Pinto et al., 2008). Several studies showing category selectivity in the higher levels of IT-tuned DNNs have used these viewpoint-controlled stimuli (Cadieu et al., 2014; Hong et al., 2016; Yamins et al., 2014). In one study, designed to evaluate different feature-based categorical coding hypotheses using these viewpoint-controlled object stimuli, Hong et al. (2016) investigated both category decoding as well as the ability of classifiers to discriminate more low-level properties of the

stimuli (e.g. object perimeter and aspect ratio) based on recordings from monkey V4 and IT. Interestingly, not only did they find that decodability for category increased between these regions (and relative to a model of early visual cortex), but also that discriminability of visual features increased across regions. These results not only disambiguate between coding hypotheses, but in addition predict that higher-level regions may code for some object-related visual features to a larger extent than one might expect intuitively.

In conclusion, we suggest that testing for residual categorical effects can distinguish between categorical and featural hypotheses. This can be done in two ways: orthogonalizing stimulus dimensions or, when this is not possible, testing multiple models with methods that allow to measure the amount of variance explained by each model separately. Crucially, as we have briefly reviewed here, where studies have made category membership and visual features orthogonal, there have indeed been residual categorical effects.

5. Converging evidence for categorical coding in the ventral pathway

In the previous section we suggested the strongest way to test for residual categorical effects with MVPA is to use stimuli where target visual features and category membership are orthogonal. However, this sort of stimulus manipulation may not always be feasible since, at the limit, one cannot control for all visual features if a categorical code is indeed feature-based. Sometimes, other properties of stimuli might rule out a featural coding hypothesis. For example, hands and tools share little in common in terms of shape or low-level visual features, so it is implausible that these features can fully explain category selectivity in the hand/tool region of LOTC (Bracci et al., 2012). A better explanation is that the region codes for their action-related properties (Bracci et al., 2016; Bracci and Peelen, 2013). However, more generally, evidence of residual categorical effects revealed with MVPA at best provides only one line of what must be converging evidence for categorical coding in a ventral pathway region.

In this section we consider three kinds of evidence that we think are especially pertinent to evaluating coding hypotheses: (i) Correlations between feature maps; (ii) deficits or disorders due to brain lesions or atypical neural development in the visual system; and (iii) short-term experimental disruption of neural functioning that impairs categorization performance.

5.1. Correlations between (diagnostic) feature maps and categorical coding

It has long been a debate whether the univariate contrasts of different categories identify parts of topographic maps for visual features in VTC, or are instead distinct spatially adjacent category selective modules (Grill-Spector and Weiner, 2014; Haxby et al., 2001; Kanwisher, 2010; Op de Beeck et al., 2008a, 2008b). For example, a consistent finding in the literature is that contrasts between faces and places isolate FFA and PPA in adjacent regions of VTC (Downing et al., 2006). However, these regions also show contrasting response biases to stimuli with particular visual features.

First, voxels in FFA show preferential responses to images with curved shapes (Caldara et al., 2006; Caldera and Seghier, 2009), low spatial frequency (SF) content (Rajimehr et al., 2011; Woodhead et al., 2011), and foveal retinotopic position (Hasson et al., 2002; Levy et al., 2001). This combination of preferences is not simply inherited from low-level visual cortex—in fact, quite the contrary. In EVC, which has a robust retinotopic organization, responses to high SF stimuli decreases with increased eccentricity (Henriksson et al., 2008; Sasaki et al., 2001). FFA shows the opposite pattern: low SF is paired with a foveal preference (Fig. 4). Second, consider that PPA—which is generally activated when viewing visual scenes, landmarks, and buildings (Epstein and Kanwisher, 1998)—may be selective for rectilinear shapes and high SFs, which are abundant in scene images (Nasr and Tootell, 2012; Nasr

		Selectivity Bias			
		category	spatial frequency	eccentricity	shape
Visual Brain Areas	EVC	-	high in central low in periphery	nonlinear magnification	-
	FFA	face	low	central	curved
	PPA	scenes buildings houses furniture	high	peripheral	rectilinear

Fig. 4. Selectivity biases in regions of the ventral pathway. Portions of VTC (FFA and PPA) show reverse feature selectivity to what is observed in EVC.

et al., 2014; Rajimehr et al., 2011; though see Bryan et al., 2016), peripherally presented images (Hasson et al., 2002; Levy et al., 2001) and to large objects (Konkle and Oliva, 2012; Park et al., 2015). As with FFA, the relation between SF preference and eccentricity bias in PPA goes in the opposite direction of that observed in EVC (Fig. 3).

One might be tempted to interpret these overlapping biases in stimulus SF and retinotopic location in terms of a diagnostic featural code or even a more low-level featural code (cf. Andrews et al., 2015). However, these biases can be equally explained if we assume these regions have some functional importance related to the representation of either faces (Bracci and Op de Beeck, 2016; Op de Beeck et al., 2008a), or scenes (Groen et al., 2017). Consider in particular that biases in SF and eccentricity in FFA suggest a functional importance to face representation, since we tend to foveate and rely on low SF information when identifying and discriminating faces (Caldara et al., 2006). Other results, which show a sensitivity in face regions to diagnostic properties of stimuli, also seem hard to reconcile with a diagnostic featural coding hypothesis. For example, strong effects of gaze behavior towards canonical feature location of faces can be linked to greater decoding in occipital face area (OFA) when face features are presented in their canonical location, such as eye stimuli presented in upper quadrants of the visual field (de Haas et al., 2016). Furthermore, FFA and OFA may even exhibit a “faciotopic” organization (Henriksson et al., 2015; van der Hurk et al., 2015).

In short, the combination of feature selectivity of these regions seems to suggest an organization specialized for representing a particular object category, faces, and is consistent with the idea that the representation of object category is feature-based and depends on detecting diagnostic fragments of stimuli in canonical retinotopic positions (Kravitz et al., 2008; Ullman, 2007). VTC does not contain a multiplexing of a series of *independent* maps of visual features. Instead, it shows marked correlations between feature maps and preferences that align well with the categories that are represented in VTC and its sub-regions. The correlations between maps for visual features of stimuli support the notion of categorical coding rather than diagnostic feature coding: the correlations mostly make sense in the context of coding for category.

5.2. Neural deficits and categorical coding

A large number of category-specific deficits have been reported in the literature, and are often targeted at specific categories, including many of the categories we have discussed, including faces, animacy and tools (for review, see Capitani et al., 2003). Although issues of categorical and featural coding also arise with respect to neuropsychological findings (Caramazza and Shelton, 1998), these deficits are generally considered category-specific, in some cases include damage to ventral pathway regions, and converge with the effects of category-specificity we have described (Mahon and Caramazza, 2009).

For instance, lesions within the right fusiform gyrus selectively

impair face recognition leaving general object recognition unaffected (Barton et al., 2002; Wada and Yamamoto, 2001), and focal lesions of the right anterior temporal lobe cause face specific impairment (Busigny et al., 2014). Also, acquired prosopagnosia causes face-specific deficits in recognition, ruling out the possibility that the disorder reflects a general deficit in recognizing visually homogeneous objects (Busigny et al., 2010). Using MVPA, some studies have also investigated the neural responses of individuals with congenital prosopagnosia. Rivolta et al. (2014) found abnormal activity in different aspects of the face network in individuals with congenital prosopagnosia, and Zhang et al. (2015) found impaired representation of face configurations in right FFA.

Finally, congenital blindness is an especially important test case for evaluating categorical coding hypotheses in ventral pathway regions, since any organization revealed in the areas cannot be the result of visual experience. Crucially, it has been found that categorical organization is preserved in congenitally blind individuals in LOTC and VTC (Handjaras et al., 2016; Peelen et al., 2014; van den Hurk et al., 2017). Although, one could argue that object shape could be acquired via other modalities such as touch (Amedi et al., 2001) or sound (Arnott et al., 2013), it is impossible for image statistics to influence visual cortex organization of these individuals. Instead, categorical organization for some divisions is likely innately specified, and present even in the absence of visual experience (Mahon and Caramazza, 2011).

A recent paper from our lab illustrates these points. van den Hurk et al. (2017) found that auditory cues for different categories (faces, bodies, scenes, and other objects) produced decodable pattern responses in the VTC of congenitally blind individuals. Even more striking, when classifiers were trained on the neural patterns of the blind individuals they could reliably generalize to predict the neural patterns of sighted individuals produced by visual images of objects from these categories. While it is difficult to deduce what precisely is shared by the representations in the VTC of blind and sighted individuals, it cannot be a reflection of visual features. In contrast, these results are consistent with VTC implementing representations of category that are partially innately specified, and further augmented in development by (multi-modal) sensory inputs. Thus, VTC may implement a feature-based categorical code, but one that need not be exclusively dependent on visual features and instead also includes features from other modalities.

5.3. Experimental interventions and categorical coding

Evidence for categorical coding also comes from research that involves experimental interventions in ventral pathway regions. This is especially the case with respect to regions of the face network. In monkeys, Afraz et al. (2006) found that microstimulation of face patches biased the response of animals performing a face recognition task. Similarly, Afraz et al. (2015) used optogenetics to stimulate clusters of face-selective neurons in monkey IT, which resulted in reduced accuracy in a face gender discrimination task.

More dramatic evidence of categorical coding for faces comes from direct electrical stimulation of portions of human fusiform gyrus anterior to FFA (Jonas et al., 2015; Parvizi et al., 2012). These studies show that when directly stimulated this region produces transient prosopagnosia in individuals' ability to recognize individual faces, causing profound face perceptual distortion, even though their ability to recognize faces as such seems to remain wholly intact. Thus one cannot explain these effects by appeal to disruption of some sort of a diagnostic featural code representation.

While these and other studies we have mentioned largely relate to one category, faces, their results highlight that there are diverse sources of evidence in favor of some form of categorical coding in the ventral pathway, which are highly relevant to the interpretation of MVPA results that have been used to support different coding hypotheses.

6. General discussion

In this paper we have discussed the debate regarding object category selectivity in the ventral visual pathway—a debate that has become particularly lively due to the analytic power provided by MVPA. As we have emphasized, it is important that one frames matters in terms of competing coding hypotheses. After comparing these hypotheses to evidence from the literature, we believe that present results generally favor feature-based categorical coding in regions of the pathway. In this concluding section we will discuss some general morals of our discussion, and suggestions for future directions.

With respect to MVPA, as we have argued stimulus design and model selection should be tailored to differentiating between diagnostic featural coding and feature-based categorical coding hypotheses. Where prominent studies may be subject to possible confounds of shape and color, it makes sense to question what of these results can be accounted for by visual features. At the same time, these sorts of results should not be taken as necessarily providing the strongest evidence in favor of categorical coding in the ventral pathway. As is true of many branches of science, evidence must converge, and the popularity of investigating object categorization in the ventral stream using MVPA provides ample opportunity for integration.

One important point of caution is that the presence of categorical information or organization in a brain region does not entail that such information or structure is exploited in a task related manner, or explicitly coded for (Cox and Savoy, 2003; de Wit et al., 2016; Ritchie et al., in press). This point holds even when showing evidence of residual categorical effects. The fact the categorical specificity in a brain region is orthogonal to target visual features of stimuli is not sufficient evidence to show that it represents object category, even if we can rule out alternative featural coding hypotheses. One way to partially address this concern is to pair MVPA with behavioral methods (Tong and Pratte, 2012). For example, Carlson et al. (2014) were able to predict reaction times of observers on an animacy task based on distances of activity patterns from a decision boundary through activation space in human IT using the data of Kriegeskorte et al. (2008a). This result suggests that the information in the region measured with fMRI could be functionally exploited, and points to a general strategy for relating psychological models of behavior to MVPA methods (Ritchie and Carlson, 2016).

While we have focused on categorical vs. featural coding in the visual system, similar issues arise in research on the neural loci of conceptual understanding. Although our discussion is orthogonal to the question of whether concepts are abstract or embodied (a distinction that may not even be fruitful for cognitive neuroscience; see Leshinskaya and Caramazza, 2016), more directly relevant is whether part of our conceptual knowledge is partially implemented in (or adjacent to) modality-specific brain regions. For example, much of the research on the neural localization of conceptual understanding posits categorical representations in the ventral pathway, and has sometimes relied on similar results to the ones we have reviewed here (Mahon and Caramazza, 2009; Martin, 2007). We believe this raises an important point: if our goal is to investigate what kind of categorical code is (or is not) found in ventral pathway regions, then we may need to broaden our perspective to consider how these regions factor into categorical representation in the brain more generally.

In sum, evidence available in the literature suggests that category selectivity cannot be accounted by feature selective biases observed throughout visual cortex. Instead, the existence of feature biases, the presence of residual categorical effects, provide strong support for the hypothesis that they all share the common goal of representing object category.

Funding

This work was supported by the European Research Council (ERC-2011-StG-284101), a federal research action (IUAP-P7/11), and a

Hercules grant ZW11_10 to H.O.P. S.B. was funded by FWO (Fonds Wetenschappelijk Onderzoek) postdoctoral fellowship (12S1317N). This project has received funding from the FWO and European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 665501, via a FWO [PEGASUS]² Marie Skłodowska-Curie fellowship (12T9217N) to J.B.R.

References

- Afraz, A., Boyden, E.S., DiCarlo, J.J., 2015. Optogenetic and pharmacological suppression of spatial clusters of face neurons reveal their causal role in face gender discrimination. *Proc. Natl. Acad. Sci. USA* 112, 6730–6735.
- Afraz, S., Kiani, R., Esteky, H., 2006. Microstimulation of inferotemporal cortex influences face categorization. *Nature* 442, 692–695.
- Amedi, A., Malach, R., Hendler, T., Peled, S., Zohary, E., 2001. Visuo-haptic object-related activation in the ventral visual pathway. *Nat. Neurosci.* 4, 324–330.
- 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. *J. Vision* (15,1-12,3).
- Arnott, S.R., Thaler, L., Milne, J.L., Kish, D., Goodale, M.A., 2013. Shape-specific activation of occipital cortex in an early blind echolocation expert. *Neuropsychologia* 51, 938–949.
- Baeck, A., Kravitz, D., Baker, C., Op de Beeck, H., 2015. Influence of lexical status and orthographic similarity on the mult[HYPHEN]voxel response of the visual word form area. *NeuroImage* 111, 321–328.
- Baldassi, C., Alemi-Neissi, A., Pagan, M., Dicarlo, J.J., Zecchina, R., Zoccolan, D., 2013. Shape similarity, better than semantic membership, accounts for the structure of visual object representations in a population of monkey inferotemporal neurons. *PLoS Comput. Biol.* 9, e1003167.
- Barton, J.J.S., Press, D.Z., Keenan, J.P., O'Connor, M., 2002. Lesions of the fusiform face area impair perception of facial configuration in prosopagnosia. *Neurology* 58, 71–78.
- Biederman, I., 2000. Recognizing depth-rotated objects: a review of recent research and theory. *Spat. Vision* 13, 241–253.
- Blanz, V., Tarr, M.J., Bülthoff, H.H., 1999. What object attributes determine canonical views? *Perception* 28, 575–599.
- Bracci, S., Caramazza, A., Peelen, M.V., 2015. Representational similarity of body parts in human occipitotemporal cortex. *J. Neurosci.* 35, 12977–12985.
- Bracci, S., Cavina-Pratesi, C., Connolly, J.D., Ietswaart, M., 2016. Representational content of occipitotemporal and parietal tool areas. *Neuropsychologia* 84, 81–88.
- Bracci, S., Cavina-Pratesi, C., Ietswaart, M., Caramazza, A., Peelen, M.V., 2012. Closely overlapping responses to tools and hands in left lateral occipitotemporal cortex. *J. Neurophysiol.* 107, 1443–1456.
- Bracci, S., Op de Beeck, H., 2016. Dissociations and associations between shape and category representations in the two visual pathways. *J. Neurosci.* 36, 432–444.
- Bracci, S., Peelen, M.V., 2013. Body and object effectors: the organization of object representations in high-level visual cortex reflects body-object interactions. *J. Neurosci.* 33, 18247–18258.
- Brincat, S.L., Connor, C.E., 2004. Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nat. Neurosci.* 7, 880–886.
- Bryan, P.B., Julian, J.B., Epstein, R.A., 2016. Rectilinear edge selectivity is insufficient to explain the category selectivity of the parahippocampal place area. *Front. Hum. Neurosci.* 10, 1–12.
- Busigny, T., Graf, M., Mayer, E., Rossion, B., 2010. Acquired prosopagnosia as a face specific disorder: ruling out the general visual similarity account. *Neuropsychologia* 48, 2051–2067.
- Busigny, T., Van Belle, G., Jemel, B., Hosen, A., Joubert, S., Rossion, B., 2014. Face-specific impairment in holistic perception following focal lesion of the right anterior temporal lobe. *Neuropsychologia* 56, 312–333.
- Cadieu, C.F., Hong, H., Yamins, D.L., Pinto, N., Ardila, D., Solomon, E.A., Majaj, N.J., DiCarlo, J.J., 2014. Deep neural networks rival the representation of primate IT cortex for core visual object recognition. *PLoS Comput. Biol.* 10, e1003963.
- Caldara, R., Seghier, M.L., 2009. The fusiform face area responds automatically to statistical regularities optimal for face categorization. *Hum. Brain Mapp.* 30, 1615–1625.
- Caldara, R., Seghier, M.L., Rossion, B., Lazeyras, F., Michel, C., Hauert, C.A., 2006. The fusiform face area is tuned for curvilinear patterns with more high-contrast elements in the upper part. *NeuroImage* 31, 313–319.
- Capitani, E., Laiacona, M., Mahon, B., Caramazza, A., 2003. What are the facts of semantic category-specific deficits? A critical review of the clinical evidence. *Cogn. Neuropsychol.* 20, 213–261.
- Caramazza, A., Shelton, J.R., 1998. Domain-specific knowledge systems in the brain: the animate-inanimate distinction. *J. Cogn. Neurosci.* 10, 1–34.
- Carlson, T.A., Schrater, P., He, S., 2003. Patterns of activity in the categorical representations of objects. *J. Cogn. Neurosci.* 15, 704–717.
- Carlson, T.A., Ritchie, J.B., Kriegeskorte, N., Durvasula, S., Ma, J., 2014. Reaction time for object categorization is predicted by representational distance. *J. Cogn. Neurosci.* 26, 132–142.
- Cichy, R.M., Pantazis, D., Oliva, A., 2014. Resolving human object recognition in space and time. *Nat. Neurosci.* 17, 455–462.
- Cichy, R.M., Pantazis, D., Oliva, A., 2016. Similarity-based fusion of MEG and fMRI reveals spatio-temporal dynamics in human cortex during visual object recognition. *Cereb. Cortex* 26, 3563–3579.

- Clarke, A., Tyler, L.K., 2014. Object-specific semantic coding in human perirhinal cortex. *J. Neurosci.* 34, 4766–4775.
- 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.
- Cohen, L., Dehaene, S., Naccache, L., Lehericy, S., Dehaene-Lambertz, G., Henaff, 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* 123, 291–307.
- Connolly, A.C., Guntupalli, J.S., Gors, J., Hanke, M., Halchenko, Y.O., Wu, Y., Abdi, H., Haxby, J.V., 2012. Representation of biological classes in the human brain. *J. Neurosci.* 32, 2608–2618.
- Connolly, A.C., Sha, L., Guntupalli, J.S., Oosterhof, N., Halchenko, Y.O., Nastase, S.A., Visconti di Oleggio Castello, M., Abdi, H., Jobst, B.C., Gobbini, M.I., Haxby, J.V., 2016. How the human brain represents perceived dangerousness or “predacity” of animals. *J. Neurosci.* 36, 5373–5384.
- Cox, D., 2014. Do we understand high-level vision? *Curr. Opin. Neurobiol.* 25, 187–193.
- 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, 261–270.
- Cutzu, F., Edelman, S., 1994. Canonical views in object representation and recognition. *Vis. Res.* 34, 3037–3056.
- Davis, T., LaRocque, K.F., Mumford, J.A., Norman, K.A., Wagner, A.D., Poldrack, R.A., 2014. What do differences between multi-voxel and univariate analysis mean? How subject-, voxel-, and trial-level variance impact fMRI analysis. *NeuroImage* 97, 271–283.
- de Wit, L., Alexander, D., Vebjorn, E., Wagemans, J., 2016. Is neuroimaging measuring information in the brain? *Psychon. Bull. Rev.* <http://dx.doi.org/10.3758/s13423-016-1002-0>.
- DiCarlo, J., Cox, D., 2007. Untangling invariant object recognition. *Trends Cogn. Sci.* 11, 333–341.
- DiCarlo, J., Zoccolan, D., Rust, N., 2012. How does the brain solve visual object recognition? *Neuron* 73, 415–434.
- Downing, P.E., Jiang, Y., Shuman, M., Kanwisher, N., 2001. A cortical area selective for visual processing of the human body. *Science* 293, 2470–2473.
- Downing, P.E., Chan, A.Y., Peelen, M.V., Dodds, C.M., Kanwisher, N., 2006. Domain specificity in visual cortex. *Cereb. Cortex* 16, 1453–1461.
- Dubois, J., de Berker, A.O., Tsao, D.Y., 2015. Single-unit recordings in the macaque face patch system reveal limitations of fMRI MVPA. *J. Neurosci.* 35, 2791–2802.
- Edelman, S., Grill-Spector, K., Kushnir, T., Malach, R., 1998. Toward direct visualization of the internal shape representation space by fMRI. *Psychobiology* 26, 309–321.
- Epstein, R., Kanwisher, N., 1998. A cortical representation of the local visual environment. *Nature* 392, 598–601.
- Grill-Spector, K., Weiner, K.S., 2014. The functional architecture of the ventral temporal cortex and its role in categorization. *Nat. Rev. Neurosci.* 15, 536–548.
- Groen, I.L., Silson, E.H., Baker, C.I., 2017. Contributions of low-and high-level properties to neural processing of visual scenes in the human brain. *Philos. Trans. R. Soc. B* 372, 20160102.
- Gross, C.G., 2008. Single neuron studies of inferior temporal cortex. *Neuropsychologia* 46, 841–852.
- Gross, C.G., Bender, D.B., Rocha-Miranda, C.E., 1969. Visual receptive fields of neurons in inferotemporal cortex of the monkey. *Science* 166, 1303–1306.
- Gross, C.G., Roche-Miranda, G.E., Bender, D.B., 1972. Visual properties of neurons in the inferotemporal cortex of the macaque. *J. Neurophysiol.* 35, 96–111.
- de Haas, B., Schwarzkopf, D.S., Alvarez, I., Lawson, R.P., Henriksen, L., Kriegeskorte, N., Rees, G., 2016. Perception and processing of faces in the human brain is tuned to typical feature locations. *J. Neurosci.* 36, 9289–9302.
- Handjaras, G., Ricciardi, E., Leo, A., Lenci, A., Cecchetti, L., Cosottini, M., Marotta, G., Pietrini, P., 2016. How concepts are encoded in the human brain: a modality independent, category-based cortical organization of semantic knowledge. *NeuroImage* 135, 232–242.
- 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, 156–166.
- Hasson, U., Levy, I., Behrmann, M., Hendler, T., Malach, R., 2002. Eccentricity bias as an organizing principle for human high-order object areas. *Neuron* 34, 479–490.
- Haxby, J.V., Connolly, A.C., Guntupalli, J.S., 2014. Decoding neural representational spaces using multivariate pattern analysis. *Annu. Rev. Neurosci.* 37, 435–456.
- Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P., 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293, 2425–2430.
- Haynes, J.D., Rees, G., 2005. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nat. Neurosci.* 8, 686–691.
- Hayward, W.G., 2003. After the viewpoint debate: where next in object recognition? *Trends Cogn. Sci.* 7, 425–427.
- Henriksson, L., Mur, M., Kriegeskorte, N., 2015. Faciotopy—a face-feature map with face-like topology in the human occipital face area. *Cortex* 72, 156–167.
- Henriksson, L., Nurminen, L., Hyvärinen, A., Vanni, S., 2008. Spatial frequency tuning in retinotopic visual areas. *J. Vision.* 8 (1–13), 5.
- Harnad, S., 1987. Psychophysical and cognitive aspects of categorical perception: a critical overview. In: Harnad, S. (Ed.), *Categorical Perception: The Groundwork of Cognition*. Cambridge University Press, pp. 1–52.
- Hong, H., Yamins, D.L., Majaj, N.J., DiCarlo, J.J., 2016. Explicit information for category-orthogonal object properties increases along the ventral stream. *Nat. Neurosci.* 19, 613–622.
- Hung, C., Kreiman, G., Poggio, T., DiCarlo, J., 2005. Fast read-out of object identity from macaque inferior temporal cortex. *Science* 310, 863–866.
- van den Hurk, J., van Baelen, M., Op de Beeck, H.P., 2017. Development of visual category selectivity in ventral visual cortex does not require visual experience. *Proc. Natl. Acad. Sci. USA*.
- van den Hurk, J., Pegado, F., Martens, F., de Beeck, H.P.O., 2015. The search for the face of the visual homunculus. *Trends Cogn. Sci.* 19, 638–641.
- Jordan, M.C., Greene, M.R., Beck, D.M., Fei-Fei, L., 2015. Basic-level category structure emerges gradually across human occipito-temporal cortex. *J. Cogn. Neurosci.* 27, 1427–1446.
- Ishai, A., Ungerleider, L.G., Martin, A., Schouten, J.L., Haxby, J.V., 1999. Distributed representation of objects in the human ventral visual pathway. *Proc. Natl. Acad. Sci. USA* 96, 9379–9384.
- Issa, E.B., Papanastassiou, A.M., DiCarlo, J.J., 2013. Large-scale, high-resolution neurophysiological maps underlying fMRI of macaque temporal lobe. *J. Neurosci.* 33, 15207–15219.
- Jonas, J., Rossion, B., Brissart, H., Frismand, S., Jacques, C., Hossu, G., Colnat-Coulbois, S., Vespignani, H., Vignal, J.P., Maillard, L., 2015. Beyond the core face-processing network: intracerebral stimulation of a face-selective area in the right anterior fusiform gyrus elicits transient prosopagnosia. *Cortex* 72, 140–155.
- Jozwik, K.M., Kriegeskorte, N., Mur, M., 2016. Visual features as stepping stones toward semantics: explaining object similarity in IT and perception with non-negative least squares. *Neuropsychologia* 83, 201–226.
- Kamitani, Y., Tong, F., 2005. Decoding the visual and subjective contents of the human brain. *Nat. Neurosci.* 8, 679–685.
- Kanwisher, N., 2010. Functional specificity in the human brain: a window into the functional architecture of the mind. *Proc. Nat. Acad. Sci. USA*, 107, pp. 11163–11170.
- Kanwisher, N., McDermott, J., Chun, M.M., 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* 17, 4302–4311.
- Kaiser, D., Azzalini, D.C., Peelen, M.V., 2016. Shape-independent object category responses revealed by MEG and fMRI decoding. *J. Neurophysiol.* 115, 2246–2250.
- Khaligh-Razavi, S.M., Kriegeskorte, N., 2014. Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS Comput. Biol.* 10, e1003915.
- Kiani, R., Esteky, H., Mirpour, K., Tanaka, K., 2007. Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *J. Neurophysiol.* 97, 4296–4309.
- Konkle, T., Oliva, A., 2012. A real-world size organization of object responses in occipitotemporal cortex. *Neuron* 74, 1114–1124.
- Kravitz, D.J., Saleem, K.S., Baker, C.I., Ungerleider, L.G., Mishkin, M., 2013. The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends Cogn. Sci.* 17, 26–49.
- Kravitz, D.J., Vinson, L.D., Baker, C.I., 2008. How position dependent is visual object recognition? *Trends Cogn. Sci.* 12, 114–122.
- Kriegeskorte, N., 2015. Deep neural networks: a new framework for modeling biological vision and brain information processing. *Ann. Rev. Vis. Sci.* 1, 417–446.
- Kriegeskorte, N., Bandettini, P., 2007. Analyzing for information, not activation, to exploit high-resolution fMRI. *NeuroImage* 38, 649–662.
- Kriegeskorte, N., Diedrichsen, J., 2016. Inferring brain-computational mechanisms with models of activity measurements. *Philos. Trans. R. Soc. B* 371, 20160278.
- Kriegeskorte, N., Mur, M., Bandettini, P., 2008a. Representational similarity analysis—connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* 24.
- Kriegeskorte, N., Mur, M., Ruff, D.A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., Bandettini, P.A., 2008b. Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* 60, 1126–1141.
- Kriegeskorte, N., Kievit, R.A., 2013. Representational geometry: integrating cognition, computation, and the brain. *Trends Cogn. Sci.* 17, 401–412.
- Kubilius, J., Bracci, S., de Beeck, H.P.O., 2016. Deep neural networks as a computational model for human shape sensitivity. *PLoS Comput. Biol.* 12, e1004896.
- Leshinskaya, A., Caramazza, A., 2016. For a cognitive neuroscience of concepts: moving beyond the grounding issue. *Psychon. Bull. Rev.* 23, 991–1001.
- Levy, I., Hasson, U., Avidan, G., Hendler, T., Malach, R., 2001. Center-periphery organization of human object areas. *Nat. Neurosci.* 4, 533–539.
- Liu, H., Agam, Y., Madsen, J.R., Kreiman, G., 2009. Timing, timing, timing: fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron* 62, 281–290.
- Martin, A., 2007. The representation of object concepts in the brain. *Annu. Rev. Psychol.* 58, 25–45.
- Marr, D., 1982. *Vision*. W.H. Freeman, San Francisco.
- Mahon, B.Z., 2015. What is embodied about cognition? *Lang. Cogn. Neurosci.* 30, 420–429.
- Mahon, B.Z., Caramazza, A., 2008. A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *J. Physiol. -Par.* 102, 59–70.
- Mahon, B.Z., Caramazza, A., 2009. Concepts and categories: a cognitive neuropsychological perspective. *Ann. Rev. Psychol.* 60, 27–51.
- Mahon, B.Z., Caramazza, A., 2011. What drives the organization of object knowledge in the brain? *Trends Cogn. Sci.* 15, 97–103.
- Malone, P.S., Glezer, L.S., Kim, J., Jiang, X., Riesenhuber, M., 2016. Multivariate pattern analysis reveals category-related organization of semantic representations in anterior temporal cortex. *J. Neurosci.* 36, 10089–10096.
- Mishkin, M., Ungerleider, L.G., Macko, K.A., 1983. Object vision and spatial vision: two cortical pathways. *Trends Neurosci.* 6, 414–417.
- Nasr, S., Echarvarria, C.E., Tootell, R.B.H., 2014. Thinking outside the box: rectilinear shapes selectively activate scene-selective cortex. *J. Neurosci.* 34, 6721–6735.
- Nasr, S., Tootell, R.B.H., 2012. A cardinal orientation bias in scene-selective visual cortex. *J. Neurosci.* 32, 14921–14926.

- O'Toole, A.J., Jiang, F., Abdi, H., Haxby, J.V., 2005. Partially distributed representations of objects and faces in ventral temporal cortex. *J. Cogn. Neurosci.* 17, 580–590.
- Oliva, A., Torralba, A., 2001. Modeling the shape of the scene: a holistic representation of the spatial envelope. *IJCV* 42, 145–175.
- Op de Beeck, H., 2010. Against hyperacuity in brain reading: spatial smoothing does not hurt multivariate fMRI analyses? *Neuroimage* 49, 1943–1948.
- Op de Beeck, H.P., Brants, M., Baeck, A., Wagemans, J., 2010. Distributed subordinate specificity for bodies, faces, and buildings in human ventral visual cortex. *Neuroimage* 49, 3414–3425.
- Op de Beeck, H.P., Haushofer, J., Kanwisher, N.G., 2008a. Interpreting fMRI data: maps, modules and dimensions. *Nat. Rev. Neurosci.* 9, 123–135.
- Op de Beeck, H.P., Torfs, K., Wagemans, J., 2008b. Perceived shape similarity among unfamiliar objects and the organization of the human object vision pathway. *J. Neurosci.* 28, 10111–10123.
- Op de Beeck, H., Wagemans, J., Vogels, R., 2001. Inferotemporal neurons represent low-dimensional configurations of parameterized shapes. *Nat. Neurosci.* 4, 1244–1252.
- Op de Beeck, H., Wagemans, J., Vogels, R., 2003. The effect of category learning on the representation of shape: dimensions can be biased but not differentiated. *J. Exp. Psychol. -Gen.* 132, 491–511.
- Palmeri, T.J., Gauthier, I., 2004. Visual object understanding. *Nat. Rev. Neurosci.* 5 (4), 291–303.
- Panzeri, S., Macke, J.H., Gross, J., Kayser, C., 2015. Neural population coding: combining insights from microscopic and mass signals. *Trends Cogn. Sci.* 19, 162–172.
- Park, S., Konkle, T., Oliva, A., 2015. Parametric coding of the size and clutter of natural scenes in the human brain. *Cereb. Cortex* 25, 1792–1805.
- Parvizi, J., Jacques, C., Foster, B.L., Witthoft, N., Rangarajan, V., Weiner, K.S., Grill-Spector, K., 2012. Electrical stimulation of human fusiform face-selective regions distorts face perception. *J. Neurosci.* 32, 14915–14920.
- Peelen, M.V., Caramazza, A., 2012. Conceptual object representations in human anterior temporal cortex. *J. Neurosci.* 32, 15728–15736.
- Peelen, M.V., He, C., Han, Z., Caramazza, A., Bi, Y., 2014. Nonvisual and visual object shape representations in occipitotemporal cortex: evidence from congenitally blind and sighted adults. *J. Neurosci.* 34, 163–170.
- Peissig, J.J., Tarr, M.J., 2007. Visual object recognition: do we know more now than we did 20 years ago? *Annu. Rev. Psychol.* 58, 75–96.
- Pinsk, M.A., Arcaro, M., Weiner, K., Kalkus, J., Inati, S., Gross, C.G., Kastner, S., 2009. Neural representations of faces and body parts in macaque and human cortex: a comparative fMRI study. *J. Neurophysiol.* 101, 2581–2600.
- Pinto, N., Cox, D.D., DiCarlo, J.J., 2008. Why is real-world visual object recognition hard? *PLoS Comput. Biol.* 4, e27.
- Pouget, A., Dayan, P., Zemel, R., 2000. Information processing with population codes. *Nat. Rev. Neurosci.* 1, 125–132.
- Proklova, D., Kaiser, D., Peelen, M.V., 2016. Disentangling representations of object shape and object category in human visual cortex: the animate–inanimate distinction. *J. Cogn. Neurosci.* 28, 680–692.
- 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 Biol.* 9, e1000608. <http://dx.doi.org/10.1371/journal.pbio.1000608>.
- Rice, G.E., Watson, D.M., Hartley, T., Andrews, T.J., 2014. Low-level image properties of visual objects predict patterns of neural response across category selective regions of the ventral visual pathway. *J. Neurosci.* 34, 8837–8844.
- Ritchie, J.B., Carlson, T.A., 2016. Neural decoding and “inner” psychophysics: a distance-to-bound approach for linking mind, brain, and behavior. *Front. Neurosci.* 10, 190.
- Ritchie, J.B., Kaplan, D.M., Klein, C., In press. Decoding the brain: neural representation and the limits of multivariate pattern analysis in cognitive neuroscience. *Brit. J. Phil. Sci.* (<https://doi.org/10.1101/127233>).
- Rivolto, D., Woolgar, A., Palermo, R., Butko, M., Schmalzl, L., Williams, M.A., 2014. Multi-voxel pattern analysis (MVPA) reveals abnormal fMRI activity in both the “core” and “extended” face network in congenital prosopagnosia. *Front. Hum. Neurosci.* <http://dx.doi.org/10.3389/fnhum.2014.00925>.
- Sasaki, Y., Hadjikhani, N., Fischl, B., Liu, A.K., Marret, S., Dale, A.M., Tootell, R.B.H., 2001. Local and global attention are mapped retinotopically in human occipital cortex. *Proc. Natl. Acad. Sci. USA* 98, 2077–2082.
- Sha, L., Haxby, J.V., Abdi, H., Guntupalli, J.S., Oosterhof, N.N., Halchenko, Y.O., Connolly, A.C., 2015. The animacy continuum in the human ventral vision pathway. *J. Cogn. Neurosci.* 27, 665–678.
- Spiridon, M., Kanwisher, N., 2002. How distributed is visual category information in human occipito-temporal cortex? An fMRI study. *Neuron* 35, 1157–1165.
- Talebi, V., Baker, C.L., 2012. Natural versus synthetic stimuli for estimating receptive field models: a comparison of predictive robustness. *J. Neurosci.* 32, 1560–1576.
- Tanaka, K., Saito, H., Fukada, Y., Moriya, M., 1991. Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *J. Neurophysiol.* 66, 170–189.
- Taylor, K.I., Devereux, B.J., Acres, K., Randall, B., Tyler, L.K., 2012. Contrasting effects of feature-based statistics on the categorisation and basic-level identification of visual objects. *Cognition* 122, 363–374.
- Tong, F., Pratte, M.S., 2012. Decoding patterns of human brain activity. *Annu. Rev. Psychol.* 63, 483–509.
- Ullman, S., 2007. Object recognition and segmentation by a fragment-based hierarchy. *Trends Cogn. Sci.* 11, 58–64.
- Wada, Y., Yamamoto, T., 2001. Selective impairment of facial recognition due to ahaematoma restricted to the right fusiform and lateral occipital region. *J. Neurol. Neurosurg. Psychiatry* 71, 254–257.
- Wardle, S.G., Ritchie, J.B., 2014. Can object category selectivity in the ventral visual pathway be explained by sensitivity to low-level image properties? *J. Neurosci.* 34, 14817–14819.
- 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., 2016a. Patterns of neural response in scene-selective regions of the human brain are affected by low-level manipulations of spatial frequency. *NeuroImage* 124, 107–117.
- Watson, D.M., Young, A.W., Andrews, T.J., 2016b. Spatial properties of objects predict patterns of neural response in the ventral visual pathway. *NeuroImage* 126, 173–183.
- Woodhead, Z.V., Brownsett, S.L., Dhanjal, N.S., Beckmann, C., Wise, R.J., 2011. The visual word form system in context. *J. Neurosci.* 31, 193–199.
- Yamane, Y., Carlson, E.T., Bowman, K.C., Wang, Z., Connor, C.E., 2008. A neural code for three-dimensional object shape in macaque inferotemporal cortex. *Nat. Neurosci.* 11, 1352–1360.
- Yamins, D.L., Hong, H., Cadieu, C.F., Solomon, E.A., Seibert, D., DiCarlo, J.J., 2014. Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc. Natl. Acad. Sci. USA*, 111, pp. 8619–8624.
- Zhang, J., Liu, J., Xu, Y., 2015. Neural decoding reveals impaired face configural processing in the right fusiform face area of individuals with developmental prosopagnosia. *J. Neurosci.* 35, 1539–1548.