Journal Club

Editor's Note: These short, critical reviews of recent papers in the *Journal*, written exclusively by graduate students or postdoctoral fellows, are intended to summarize the important findings of the paper and provide additional insight and commentary. For more information on the format and purpose of the Journal Club, please see http://www.jneurosci.org/misc/ifa_features.shtml.

Can Object Category-Selectivity in the Ventral Visual Pathway Be Explained by Sensitivity to Low-Level Image Properties?

Susan G. Wardle and J. Brendan Ritchie
Perception in Action Research Centre and Department of Cognitive Science, Macquarie University, Sydney, NSW 2109, Australia
Review of Rice et al.

While object recognition typically feels effortless, it is one of the most computationally impressive feats performed by the human visual system. Due to its importance as an end stage of visual processing, a great deal of research has focused on characterizing those regions of the brain responsible for representing object categories. Congruent with research in monkeys, regions of the ventral visual pathway in humans are category-selective. Compared with the large number of possible object categories, the number of anatomically distinct regions with specific category selectivity is limited. Recent advances in neuroimaging methods such as multi-voxel pattern analysis (MVPA) have revealed a more complex and distributed neural architecture of category selectivity in the ventral stream (Haxby et al., 2001; Carlson et al., 2003; Kriegeskorte et al., 2008). For example, distributed category representations have been identified for inanimate versus animate objects (Kriegeskorte et al., 2008) and for real-world object size (Konkle and Oliva, 2012). In particular, human inferior temporal cortex (IT) has a clear categorical organization, which mirrors the organization of monkey IT (Kriegeskorte et al., 2008).

In contrast to the category representations in higher visual areas, early visual cortex responds to low-level visual properties such as orientation, spatial frequency, and luminance contrast. It is still unknown how category-selectivity emerges from tuning to low-level features earlier in the ventral stream (DiCarlo et al., 2012). An important question is to what extent the categorical organization of the ventral visual stream can be accounted for by differences in low-level image properties common to different object categories. Exemplars of the same category usually share low-level properties as well as higher-order category membership. For example, house exemplars typically include edges at certain orientations (horizontal and vertical), and face exemplars have a typical shape (round) with features at predictable positions (e.g., the relative position of the eyes and mouth). Thus, it is possible that what appears to be categorical organization of ventral cortex may alternatively be explained by the image features characteristic of object categories. However, some previous studies have ruled out selectivity for low-level image properties by demonstrating that categorical structure does not emerge when computational models of early visual processing are applied to the neuroimaging data from ventral areas

(e.g., Kriegeskorte et al., 2008). A recent paper in *The Journal of Neuroscience* by Rice et al. (2014) takes a new approach to the question, aiming to directly compare brain activity elicited by different categories of images with the low-level image statistics of each category. The authors reasoned that a correlation between low-level image statistics and patterns of neural activity would provide evidence that category selectivity in the ventral stream is better explained by sensitivity to low-level image properties.

In their primary experiment, Rice et al. (2014) used functional magnetic resonance imaging (fMRI) to measure the blood-oxygen level-dependent (BOLD) response of participants while they viewed sets of images from five categories (bottles, chairs, shoes, houses, and faces). BOLD was measured in 17 anatomically defined regions of interest (ROI). These included superior, medial, and inferior temporal regions, along with fusiform and parahippocampal areas. The neural response to each category (summed across within-category exemplars) was determined using a variant of correlation-based MVPA. Although MVPA is usually applied as a within-subjects analysis, Rice et al. (2014) added a novel twist and adapted the method of Haxby et al. (2001) to compare voxel-based patterns across subjects. The advantage of their approach is that it examines correlations for patterns of object-related activity that generalize across

subjects, instead of those specific to an individual brain. Between-subject correlations for each category were obtained using a "leave one participant out" method: the group pattern for all subjects was compared with the data for one individual excluded from the group analysis, and this procedure was repeated so that each participant was left out once. The similarity of the patterns of voxel-based activation in response to different categories was then calculated using Pearson correlation. This produced a correlation matrix comparing the patterns of fMRI response for each object category with every other category (Rice et al., 2014, their Fig. 6A).

The authors used Oliva and Torralba's (2001) computational GIST descriptor of visual scenes as a measure of the global low-level properties of the stimuli used in the fMRI experiment. The GIST model applies Gabor filters at a range of orientations and spatial frequencies to characterize the low-level image statistics across different locations in an image. After calculating the GIST descriptor for each individual exemplar image, Rice et al. (2014) computed both within- and between-category correlations. The within-category correlations were based on a leave-one-out design: the average GIST descriptor was calculated for all exemplars within a category except one, and then the remaining image was correlated with its category average. The between-category correlations were evaluated by comparing the GIST descriptor from each image with the average GIST descriptor for the images in another category. These correlations were then averaged to produce a single coefficient for each within- and between-category comparison. As would be expected, the within-category correlations were generally higher than the between-category, and this confirmed that the GIST descriptor captured some low-level image properties characteristic of each category (Rice et al., 2014, their Fig. 6B).

To assess whether low-level image properties could account for differences in the BOLD response patterns produced by images of different categories, the authors asked whether the average correlations for the GIST descriptors for each pair of categories were correlated with the average correlations for the BOLD responses to those categories. When within-category correlations were included, Rice et al. (2014) found a strong correlation between average patterns of fMRI activity and the average Fourier power spectra quantified by the GIST model for each category in the entire ventral visual cortex, which was

defined as the composite of the 17 anatomical ROIs (Rice et al., 2014, their Fig. 6C). Significant correlations were also observed individually for 10/17 of the ROIs (Rice et al., 2014, their Fig. 7, note the differences in *y*-axis scale). When within-category correlations were removed, the overall correlation was reduced and only 5/17 anatomical ROIs showed a significant correlation, including portions of the inferior temporal gyrus. Notably, the fMRI response in anterior IT was not correlated with image statistics (even when within-category correlations were included), which is intriguing as there is considerable evidence for object-selectivity in this region in both humans and monkeys (Kriegeskorte et al., 2008; Baldassi et al., 2013). This negative result suggests that object-selectivity in anterior IT cannot be reduced to sensitivity to the low-level image statistics characteristic of particular categories.

The most important question raised by the results of Rice et al. (2014) is whether they challenge the existence of object category-selective regions in ventral temporal cortex. Because some of the strongest evidence in favor of the categorical organization of human IT was provided by Kriegeskorte et al. (2008), it is useful to compare the results of Rice et al. (2014) with the results of that study. Similar to Rice et al. (2014), Kriegeskorte et al. (2008) also measured patterns of brain activity with fMRI in response to images from different categories, including animal and human faces and bodies, natural objects, and human artifacts. However, instead of averaging patterns of brain activity over experimental blocks containing different exemplars within a predefined category, Kriegeskorte et al. (2008) constructed representational dissimilarity matrices for human IT and early visual cortex using an event-related fMRI design. Representational dissimilarity matrices consist of correlations defined by every pairwise comparison of the voxel-based patterns of brain activity between the individual exemplar stimuli. A virtue of this form of analysis is that it treats each individual exemplar as its own condition and does not presuppose a categorical grouping of exemplars in the analysis. Kriegeskorte et al. (2008) found that a clear categorical structure emerged for the patterns of activity measured in human IT between animate and inanimate objects, as well as subordinate categories such as faces and bodies. Animate objects had more similar patterns of BOLD response to each other than to inanimate objects, and vice versa. Critically, this categorical

structure was not present in the dissimilarity matrix for early visual cortex. To further control for the possibility that their results could be accounted for by low-level image properties, Kriegeskorte et al. (2008, their supplemental materials) compared the dissimilarity matrix for human IT to matrices constructed from a battery of computational models of early visual processing, including a model of V1. None of these models were able to account for the categorical clustering observed in human IT.

The results of Rice et al. (2014) show that objects that have more similar low-level image statistics also tend to produce more similar patterns of brain activity as measured with fMRI. However, their results do not necessarily contradict evidence for categorical organization in ventral regions (e.g., Kriegeskorte et al., 2008). It is interesting that Rice et al. (2014) found a correlation between global image statistics and patterns of BOLD response in the occipital pole and surrounding regions of similar magnitude to that observed higher in the ventral stream. This suggests that their method of comparing image statistics and patterns of brain activity may not have been sensitive enough to detect processing differences between late and early visual areas, since there are well known functional divisions between these regions. Given the difficulty of comparing image properties and neural activity with the current sensitivity of neuroimaging techniques, the most impressive result from Rice et al. (2014) is that there is a clear correlation in the entire ventral stream between the average GIST descriptor correlations and the correlations between patterns of BOLD response when within-category correlations are removed. However, their results are more equivocal for the individual anatomical ROIs, and a lack of functional division between these areas makes it difficult to generate predictions about the relative strength of the relationship between low-level image properties and activity in each of these regions.

In summary, although it is intuitive that low-level visual features characteristic of categories must be related to category-level representations at some level, it is unlikely that they can fully account for the categorical selectivity exhibited by regions of ventral visual cortex. Moreover, Harel et al. (2014) recently demonstrated that object representations in some ventral object-selective areas are task-dependent. They used identical image sets across six tasks, and showed that

while MVPA decoding of object identity remained constant in early visual cortex between tasks, across-task decoding performance was reduced in higher object areas in the ventral stream. In conjunction with previous results that demonstrate that the categorical clusters in IT are not predicted by models of early visual processing (Kriegeskorte et al., 2008), this suggests that object category representations cannot be entirely reduced to sensitivity to low-level features. However, the results of Rice et al. (2014) are important in highlighting that information related to low-level visual properties persists even in higher visual areas. Understanding the transformation of visual information from lower to higher visual areas remains one of the major goals of visual neuroscience.

References

Baldassi C, Alemi-Neissi A, Pagan M, DiCarlo JJ, 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. CrossRef Medline

Carlson TA, Schrater P, He S (2003) Patterns of activity in the categorical representations of objects. J Cogn Neurosci 15:704–717. CrossRef Medline

DiCarlo JJ, Zoccolan D, Rust NC (2012) How does the brain solve visual object recognition? Neuron 73:415–434. CrossRef Medline

Harel A, Kravitz DJ, Baker CI (2014) Task context impacts visual object processing differentially across the cortex. Proc Natl Acad Sci U S A 111:E962–E971. CrossRef Medline

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293:2425–2430. CrossRef Medline

Konkle T, Oliva A (2012) Real-world size organization of object responses in occipitotemporal cortex. Neuron 74:1114–1124. CrossRef Medline

Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H, Tanaka K, Bandettini PA (2008) Matching categorical object representations in inferior temporal cortex of man and monkey. Neuron 60:1126–1141. CrossRef Medline

Oliva A, Torralba A (2001) Modeling the shape of the scene: a holistic representation of the spatial envelope. Int J Comput Vis 42:145–175. CrossRef

Rice GE, Watson DM, Hartley T, Andrews TJ (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. CrossRef Medline