by which the brain orchestrates region-specific dopamine signaling. Just as importantly, the finding that dopamine neuron responses track cognitive function could prove to be valuable for our understanding of Parkinson's disease, in which dopaminergic medications used for the control of motor symptoms are sometimes accompanied by cognitive side effects. Further work delineating the separate cognitive, motor, and learning signals in the SNc and VTA might eventually lead to better treatments that preferentially target dopamine's role in movement while sparing patients' cognitive abilities. Yet much remains to be done. For a long while yet, it appears, the tiny dopaminergic midbrain will continue to demand a large body of work.

REFERENCES

Berridge, K.C., and Robinson, T.E. (1998). Brain Res. Brain Res. Rev. *28*, 309–369.

Bromberg-Martin, E.S., Matsumoto, M., and Hikosaka, O. (2010). Neuron *68*, 815–834.

Fiorillo, C.D. (2013). Science *341*, 546–549.

Haber, S.N., and Knutson, B. (2010). Neuropsychopharmacology *35*, 4–26.

Horvitz, J.C. (2000). Neuroscience *96*, 651–656.

Li, B.-M., and Mei, Z.-T. (1994). Behav. Neural Biol. *62*, 134–139.

Matsumoto, M., and Hikosaka, O. (2009). Nature *459*, 837–841.

Matsumoto, M., and Takada, M. (2013). Neuron *79*, this issue, 1011–1024.

Noudoost, B., and Moore, T. (2011). Nature *474*, 372–375.

Redgrave, P., and Gurney, K. (2006). Nat. Rev. Neurosci. *7*, 967–975.

Sawaguchi, T., and Goldman-Rakic, P.S. (1991). Science *251*, 947–950.

Sawaguchi, T., and Goldman-Rakic, P.S. (1994). J. Neurophysiol. *71*, 515–528.

Schultz, W., Dayan, P., and Montague, P.R. (1997). Science *275*, 1593–1599.

Watanabe, M., Kodama, T., and Hikosaka, K. (1997). J. Neurophysiol. *78*, 2795–2798.

Williams, G.V., and Goldman-Rakic, P.S. (1995). Nature *376*, 572–575.

The Cerebral Emporium of Benevolent Knowledge

Patrick J. Mineault[1] **and Christopher C. Pack**[1,*]
[1]Montreal Neurological Institute, McGill University, Montreal, QC H3A 2B4, Canada
*Correspondence: christopher.pack@mcgill.ca
http://dx.doi.org/10.1016/j.neuron.2013.08.012

Visual objects tend to be found in predictable combinations (e.g., pens with paper). How does the brain represent these regularities? In this issue of *Neuron*, Stansbury et al. (2013) use fMRI to study the brain's representation of visual scene categories.

In a 1942 essay, Jorge Luis Borges discusses the categorization of animals, purportedly found in a fictitious Chinese encyclopedia named the "Celestial Empire of Benevolent Knowledge" (Borges, 1942). Animals therein are classified into 14 fanciful categories, including, "fabulous ones," "those that have just broken the flower vase," and "those that look like flies when viewed from a distance." Borges uses this example to suggest that any attempt to categorize the contents of nature is "arbitrary and full of conjectures."

Nevertheless (again quoting Borges), "the impossibility of penetrating the divine scheme of the universe cannot dissuade us from outlining human schemes, even though we are aware that they are provisional." In fact, such schemes can be quite useful in sensory neuroscience. A decade after Borges's essay, Barlow (1953) discovered neurons that respond selectively to stimuli that look like flies when viewed from a distance. These "fly detectors" were found in the retinas of frogs and, hence, were linked to a specific category of behavior (feeding). Subsequently, Hubel and Wiesel (1962) identified visual cortical cells that were described as "simple" and "complex," and these turned out to be useful labels for understanding many aspects of the visual cortex from anatomy to computation.

More recent imaging studies have led to the suggestion that neurons with particular stimulus selectivities are clustered together, forming brain modules responsible for encoding rather abstract categories of stimuli, including faces (Tsao et al., 2006), places (Epstein and Kanwisher, 1998), and buildings (Hasson et al., 2003). Of course, the number of such categories must be far greater than the number of brain regions, which leads to the profound question of how the brain organizes such a vast quantity of visual experience. In this issue of *Neuron*, Stansbury et al. (2013) address this question.

Stansbury et al. (2013) used fMRI imaging of human subjects to study the brain's representation of visual scene categories, defined as classes of images that contain similar co-occurrences of individual objects. For example, a scene that contains a building and a car is more likely to belong to the category "cityscape" than to the category "nautical." Obviously, one object (e.g., a tree) can be found in more than one scene (e.g., cityscape and rural), and one scene (e.g., a harbor) can belong to more than one scene category (e.g., cityscape and nautical). Thus, part of the challenge of understanding the brain's representation of scene categories is in understanding the organization of the categories themselves.

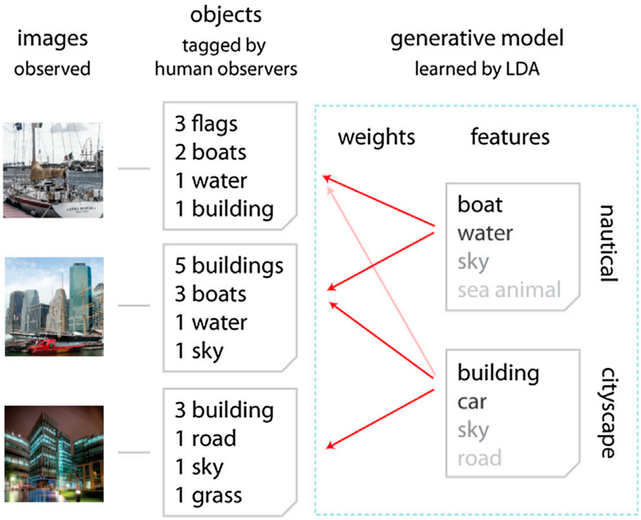

Figure 1. Estimating Categories from Natural Images
Human observers derive lists of objects from natural images (left). A generative model (right) specifies that these lists of objects are generated by weighted mixtures of features, which, in this case, are categories. The parameters of the model—the word probabilities corresponding to each category as well as the category vector corresponding to each image—are learned by the latent Dirichlet allocation algorithm.

To this end, Stansbury et al. (2013) have adopted an elegant approach that defines the scene categories objectively with an algorithm that detects the presence of certain combinations of objects in a large database of natural scenes. Importantly, the algorithm is not given any prior information about which categories each scene belongs to; it defines categories on the basis of statistical regularities. This approach largely circumvents Borges's problem of the arbitrariness of categories, given that the classification is defined by the images themselves rather than being imposed by the person doing the analysis.

In this approach, each scene (Figure 1, left) was tagged with a list of objects (e.g., two boats, one car, one person, etc.; Figure 1, middle) identified by human observers. These descriptors were fed to an unsupervised learning algorithm known as latent Dirichlet allocation (LDA), which inferred the categories represented in the data set on the basis of the pattern of co-occurrences of objects (Blei et al., 2003).

LDA, which has its root in text classification, is one of a number of unsupervised learning techniques that aim to uncover structure in complex data. Typically, they define each example in the data set—e.g., a list of words, an image, or a sound—as being generated by a noisy, weighted mixture of features. Optionally, they define a set of soft constraints, or priors, on the distribution of features and weights. The goal of the learning algorithm is to find a set of features and weights that captures the bulk of the variation in the data set while respecting the prior assumptions of the algorithm.

In LDA, each scene descriptor is assumed to be generated by a mixture of categories—the features (Figure 1, right). LDA assumes that the weights associated with this mixture (Figure 1, red arrows) are sparse—each scene contains only a handful of categories. It also assumes that weights are positive—whereas a scene may belong to a category (positive weight; indicated by a red arrow in Figure 1) or not (zero weight). It is not meaningful to say that a scene belongs negatively to a category (negative weight). The ensemble of weights linking a scene to each scene category is called the scene's category vector.

This sparse, positive encoding scheme allows the algorithm to leverage parts-based or combinatorial coding (e.g., both nautical and cityscape) in order to describe more narrowly defined scenes (e.g., harbor; Figure 1, middle). Each category is itself a sparse, positive mixture of objects (Figure 1, right).

These assumptions are embedded within a hierarchical, probabilistic model; objects contained within each category and the categories contained within each scene are jointly estimated by Bayesian inference. The resulting categories contained a high proportion of related objects. For example, one category assigned the highest weights for highway, car, sky, vehicle, and signpost—most likely corresponding to highways or ground transportation. Furthermore, the model assigned intuitive categories to the scenes in the database, tagging a harbor scene with nautical and cityscape categories. This is not surprising, given that LDA and its extensions have proven widely applicable in an analogous problem, determining categories from text documents (Blei et al., 2003).

The LDA approach taken by Stansbury et al. (2013) has revealed hidden structure in natural images, but does the visual system exploit this structure in its representation of visual scenes? One way to answer this question is to ask whether some aspect of brain activity correlates systematically with scene categories during the viewing of natural images. This would suggest that the brain encodes the scene categories in the same way that previous work has suggested an encoding of faces or orientations.

To tackle this question, Stansbury et al. (2013) had subjects view a variety of different scenes and simultaneously recorded their brain activity with fMRI. Then, the authors attempted to predict the BOLD response in each voxel under the assumption that the response to a scene was given by a weighted sum of the scene's category vector.

Responses in low-level striate and extrastriate visual areas, which are sensitive to elementary features such as orientation and contrast, were poorly modulated by scene category. However, responses in anterior visual areas such as the fusiform face area (FFA) and the parahippocampal place area (PPA) could be accurately predicted by the encoding model. The authors found that the predictions were most accurate when the LDA model contained 20 categories and 850 objects, indicating that there is substantially more categorical information available at the macroscopic fMRI scale than previously appreciated.

Importantly, the number of voxels significantly predicted by the category-encoding model was larger than alternative models relying on elementary visual features, such as orientation or spatial frequency. This was a crucial test of the hypothesis that high-level visual areas actually represent scene categories rather than visual stimuli per se (Malach et al., 1995). Consistent with this idea, the model was also significantly more accurate than others that relied only on the presence of individual objects.

Category preferences in different areas were, to some degree, consistent with previous literature. For example, the FFA showed a relative preference for the portraits category, whereas the PPA was most selective for categories that could be labeled "places." However, the results of this analysis indicated a more complex relationship between brain regions and category selectivity: voxels in several anterior visual areas showed selectivity for other categories. For example, the FFA was selective for the "plants" category in addition to "portraits." These results are consistent with earlier results from the same group, which highlighted the presence of a distributed representation of categories with smooth, overlapping gradients of preferred categories along certain cortical directions (Huth et al., 2012).

A second way to test the idea that scene categories are represented in specific brain regions is to ask whether it is possible to decode the category viewed by the observer on the basis of the BOLD activity alone. This approach is similar to that used by the same group to demonstrate how the brain represents specific images and objects (Naselaris et al., 2011). The authors found that BOLD activity successfully predicted the category membership of individual images. Importantly, these images were of novel scenes that were not used to formulate the encoding model, indicating that the model generalized beyond the specific exemplars on which it was trained.

Then, they used the LDA model to successfully predict the objects present in individual images on the basis of predicted category membership alone. This is quite a remarkable result given that objects are only encoded in the model indirectly through their correlation with scene categories. The success of this decoding approach implies that the distribution of objects in natural scenes contains substantial structure and that this structure can be exploited by the visual system.

These results might help to explain previous psychophysical findings that indicate that, when the gist of a scene is understood, objects within it can be recognized accurately even at extremely low resolutions, in some cases as low as ~6 × 6 pixels (Torralba, 2009). Performance in these tasks becomes worse when objects are isolated from their context. Similarly, human observers can detect an object more efficiently when it is found within a contextually consistent scene than when it is not (Biederman et al., 1973). Evidently, the problem of inferring object identity from low-level visual features is made much easier by context. Much like low-level how vision can make use of prior information to accurately estimate motion direction from noisy observations (Weiss et al., 2002), high-level vision could make use of learned statistical regularities to estimate object identity in ambiguous scenes (Lee and Mumford, 2003).

More generally, the approach developed by Stansbury et al. (2013) may provide an objective way to probe the brain's representation of abstract sensory information. Scene categories are abstract, in that they are largely independent of specific image features, but could they even be independent of vision? Would the sounds of traffic and the smell of baked goods produce the same activation as pictures of a city street? Perhaps sensory stimulation is not necessary at all: could imagining a specific type of scene produce interpretable activation in the relevant brain regions? Such representations might ultimately facilitate the extraction of even more abstract, perhaps semantic, information from brain activity.

REFERENCES

Barlow, H.B. (1953). J. Physiol. *119*, 69–88.

Biederman, I., Glass, A.L., and Stacy, E.W., Jr. (1973). J. Exp. Psychol. *97*, 22–27.

Blei, D.M., Ng, A.Y., and Jordan, M.I. (2003). J. Mach. Learn. Res. *3*, 993–1022.

Borges, J.L. (1942). The analytical language of John Wilkins. In Other Inquisitions (1937-1952), J.F. Solem, B.A. Davidsen, and R. Anderson, eds. (Austin, TX: University of Texas Press), pp. 101–105.

Epstein, R., and Kanwisher, N. (1998). Nature *392*, 598–601.

Hasson, U., Harel, M., Levy, I., and Malach, R. (2003). Neuron *37*, 1027–1041.

Hubel, D.H., and Wiesel, T.N. (1962). J. Physiol. *160*, 106–154.

Huth, A.G., Nishimoto, S., Vu, A.T., and Gallant, J.L. (2012). Neuron *76*, 1210–1224.

Lee, T.S., and Mumford, D. (2003). J. Opt. Soc. Am. A Opt. Image Sci. Vis. *20*, 1434–1448.

Malach, R., Reppas, J.B., Benson, R.R., Kwong, K.K., Jiang, H., Kennedy, W.A., Ledden, P.J., Brady, T.J., Rosen, B.R., and Tootell, R.B. (1995). Proc. Natl. Acad. Sci. USA *92*, 8135–8139.

Naselaris, T., Kay, K.N., Nishimoto, S., and Gallant, J.L. (2011). Neuroimage *56*, 400–410.

Stansbury, D.E., Naselaris, T., and Gallant, J.L. (2013). Neuron *79*, this issue, 1025–1034.

Torralba, A. (2009). Vis. Neurosci. *26*, 123–131.

Tsao, D.Y., Freiwald, W.A., Tootell, R.B., and Livingstone, M.S. (2006). Science *311*, 670–674.

Weiss, Y., Simoncelli, E.P., and Adelson, E.H. (2002). Nat. Neurosci. *5*, 598–604.