

Representing “stuff” in visual cortex

Corey M. Ziemba^a and Jeremy Freeman^{b,1}

^aCenter for Neural Science, New York University, New York, NY 10003; and ^bJanelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA 20147

Despite decades of study, we do not understand the fundamental processes by which our brain encodes and represents incoming visual information and uses it to guide perception and action. A wealth of evidence suggests that visual recognition is mediated by a series of areas in primate cortex known as the ventral stream, including V1 (primary visual cortex), V2, and V4 (Fig. 1A) (1). The earliest stages are to some extent understood; Hubel and Wiesel famously discovered, for example, that neurons in V1 respond selectively to the orientation and direction of a moving edge (2). However, a vast gulf remains between coding for a simple edge and representing the full richness of our visual world. David Hubel himself observed in 2012 that we still “have almost no examples of neural structures in which we know the difference between the information coming in and what is going out—what the structure is for. We have some idea of the answer for the retina, the lateral geniculate body, and the primary visual cortex, but that’s about it” (3). In PNAS, Okazawa et al. (4) make significant headway in this quest by uncovering and characterizing a unique form of neural selectivity in area V4.

A central challenge in understanding how neurons encode visual stimuli is knowing what stimulus to show the neurons. We do not know the “right” stimuli until we have some idea of what neurons are selective for, but we might not know what neurons are selective for until we have shown them the right stimuli. Near the top of the ventral stream hierarchy, such as in the inferotemporal cortex, it has proven useful to probe responses using highly complex stimuli such as photographs of natural scenes and objects (1, 5–7). However, the complexity of such stimuli and the difficulty in experimentally manipulating or controlling them can make it hard to tell what any given neuron is encoding, beyond the fact that it responds more to one picture than another. In earlier stages—the retina, lateral geniculate nucleus, and V1—the use of simple stimuli, like noise patterns, oriented edges, or sine-wave gratings, has yielded a reasonable working understanding of neural encoding (8), but such stimuli are sufficient only because these neurons exhibit simpler forms of encoding, at least to a first approximation.

In the area studied by Okazawa et al., area V4, most previous authors have characterized neurons by assuming that they encode hard-

edged shapes and contours, using stimuli stitched together from “V1-like” line segments into longer contours with parameterized curvatures (9–11). This approach reflects an intuitive understanding of the visual world: that shapes and surfaces are defined by their bounding contours and that the visual system must somehow represent these features. However, as Okazawa et al. point out, much of the visual world is characterized not by contours, but by texture: the patterns that make up the surfaces of objects and environments. Ted Adelson described this as the distinction between “things” (objects, elements of scenes) and “stuff” (materials, textures, etc.) (12).

Textures are notoriously difficult to work with as visual stimuli; unlike the angle of a line or the curvature of a contour, they do not permit a simple parameterization. What set of numbers could capture the difference between wood bark and a patch of grass? To solve this problem, Okazawa et al. drew on existing work in the modeling and synthesis of visual texture and extended it in novel ways. They began with a texture model developed by Portilla and Simoncelli (13). The model has two components: a set of statistics, computed on an image, that implicitly capture many of the higher-order properties of visual textures (Fig. 1B), and an algorithm for generating stimuli with those properties. It was originally developed to capture the perceptually relevant properties of visual texture (13, 14) and has been extended to explain how visual representations vary across the visual field (15, 16) and, qualitatively, across different ventral stream areas (6, 17).

The hundreds of parameters contained in such image statistical models have prevented their use in detailed neural characterization, but Okazawa et al. are able to transform the model into a suitable substrate for characterizing neural selectivity. First, they took a large ensemble of stimuli and used dimensionality reduction to shrink the hundreds of model parameters into a low-dimensional space. Even in this simplified space, they could not realistically show all possible parameter

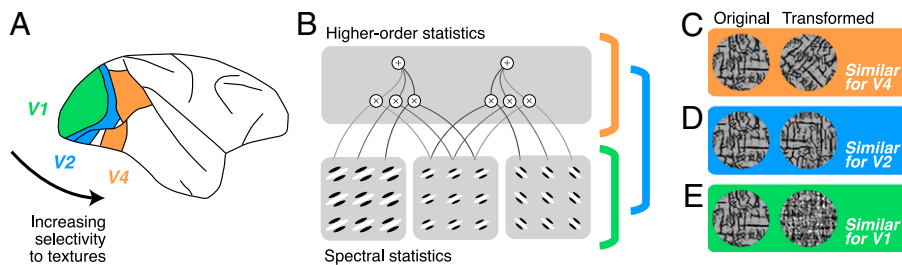


Fig. 1. (A) Lateral view of the macaque brain with early ventral stream areas V1, V2, and V4 highlighted. (B) Schematic diagram of the parameters contained in the Portilla Simoncelli texture model. Spectral statistics reflect the output of V1-like filters. Higher-order statistics reflect correlations of these filter outputs across orientations, spatial frequencies, and local positions. Most V1 neurons are only sensitive to spectral statistics, and many V2 neurons are sensitive to both spectral and higher-order statistics; Okazawa et al. show that some V4 neurons are tuned exclusively for higher-order statistics. (C–E) In early ventral stream areas, physically different images can yield similar responses, and different image transformations can reveal particular encoding properties. (C) Rotating an image changes the power spectra but preserves some higher-order statistics. In V4, as reported by Okazawa et al., differently rotated images can yield similar responses. (D) Spatially translating a texture changes the image pixel-by-pixel but preserves the spectral and higher-order statistics. In V2, such images yield similar responses. (E) Randomizing the phase of an image destroys higher-order statistics but preserves the power spectrum. In V1, images with similar spectral statistics yield similar responses, with or without higher-order statistics.

Author contributions: C.M.Z. and J.F. wrote the paper.

The authors declare no conflict of interest.

See companion article 10.1073/pnas.1415146112.

¹To whom correspondence should be addressed. Email: freemanj11@janelia.hhmi.org.

combinations to each neuron. Therefore, drawing on existing work characterizing shape selectivity (18), they used an adaptive sampling technique to explore regions of their low-dimensional space that evoked large firing rates. Having measured responses from each neuron to a sufficiently rich and response-evoking ensemble of stimuli, they could then model the response in terms of the low-dimensional space.

Previous work has examined the responses of V4 neurons to texture stimuli (6, 19), but with their modeling technique, Okazawa et al. were able to characterize in detail several largely unknown forms of selectivity in V4. First, they found that many V4 neurons were well described by selectivity to higher-order image statistics, and some were tuned to particular subsets of higher-order statistics. These subsets have curious names like “energy cross-orientation” and “linear cross-position”; although not exactly intuitive, the authors make an effort to show, with pictures, how selectivity to a particular statistic relates to preferences for particular images. Most remarkably, they use simple image manipulations to show that some V4 neurons selectively encode these higher-order statistics while remaining tolerant to changes in the “power spectra,” a term that describes the total amount of different orientations and spatial frequencies in an image, which is what V1 neurons are mostly tuned for. For example, a V4 neuron might respond well to a bark-like texture regardless of the overall orientation of the pattern (Fig. 1 B and C). Finally, by examining how well a population of V4 neurons could discriminate among different texture patterns, they were able to show that the representation of higher-order image statistics in V4 resembles the perceptual representation derived from previous behavioral experiments (17).

The modeling effort of Okazawa et al. represents both a technical advance and a conceptual counterweight to previous efforts in V4. Many models of V4 have characterized selectivity to contours, using simple parameterized feature spaces (10, 11). Rather than consider neurons as performing computations on an actual visual input, these mod-

els operate in the space of abstract quantities, like curvature, and as a result, the models only apply to those kinds of stimuli. In contrast, the current approach can make predictions about how neurons should respond to any stimulus pattern (20). In future work, the

Okazawa et al. were able to characterize in detail several largely unknown forms of selectivity in V4.

authors could use this fact to reconcile their findings with earlier results in V4. One elegant possibility is that previously described selectivity to curvature and contours is just a special case of sensitivity to higher-order image statistics. However, an alternative is that Okazawa et al. focused only on a subset of V4 cells specifically tuned to texture, whereas previous efforts described a different representation, possibly mediated by distinct but interacting neuronal populations.

The current results are also interesting in light of recent work in area V2. Whereas V1 neurons seem to encode almost exclusively

spectral properties like orientation and spatial frequency content (Fig. 1 B and E), V2 neurons additionally show selectivity for higher-order image statistics (Fig. 1 B and D), similar to those parameterized by Okazawa et al. (17). It will thus be important in the future to determine how much of the higher-order statistical selectivity in V4 is inherited from V2 or computed de novo from its inputs. However, most V2 neurons retain sensitivity to spectral properties, whereas at least some neurons in V4 appear largely tolerant to spectral changes. These findings may thus suggest a transformation from V2 to V4 that complements and extends the transformation from V1 and V2. New techniques will be required, alongside modeling and stimulus design, to characterize in mechanistic detail the computations that take place between these cortical areas and across different layers and cell types within an area (21). How exactly these cortical transformations form a physiological basis for vision remains a deeply intriguing puzzle, and approaches such as that of Okazawa et al. will help pave the way forward.

- 1 DiCarlo JJ, Zoccolan D, Rust NC (2012) How does the brain solve visual object recognition? *Neuron* 73(3):415–434.
- 2 Hubel DH, Wiesel TN (1959) Receptive fields of single neurones in the cat's striate cortex. *J Physiol* 148:574–591.
- 3 Hubel D, Wiesel T (2012) David Hubel and Torsten Wiesel. *Neuron* 75(2):182–184.
- 4 Okazawa G, Tajima S, Komatsu H (2014) Image statistics underlying natural texture selectivity of neurons in macaque V4. *Proc Natl Acad Sci USA*, 10.1073/pnas.1415146112.
- 5 Hung CP, Kreiman G, Poggio T, DiCarlo JJ (2005) Fast readout of object identity from macaque inferior temporal cortex. *Science* 310(5749):863–866.
- 6 Rust NC, Dicarlo JJ (2010) Selectivity and tolerance (“invariance”) both increase as visual information propagates from cortical area V4 to IT. *J Neurosci* 30(39):12978–12995.
- 7 Huth AG, Nishimoto S, Vu AT, Gallant JL (2012) A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron* 76(6):1210–1224.
- 8 Carandini M, et al. (2005) Do we know what the early visual system does? *J Neurosci* 25(46):10577–10597.
- 9 Gallant JL, Braun J, Van Essen DC (1993) Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. *Science* 259(5091):100–103.
- 10 Pasupathy A, Connor CE (2001) Shape representation in area V4: position-specific tuning for boundary conformation. *J Neurophysiol* 86(5):2505–2519.
- 11 Nandy AS, Sharpee TO, Reynolds JH, Mitchell JF (2013) The fine structure of shape tuning in area V4. *Neuron* 78(6):1102–1115.
- 12 Adelson EH (2001) On seeing stuff: The perception of materials by humans and machines. *Proceedings of the SPIE* 4299:1–12.
- 13 Portilla J, Simoncelli E (2000) A parametric texture model based on joint statistics of complex wavelet coefficients. *Int J Comput Vis* 40(1):49–70.
- 14 Julesz B (1962) Visual pattern discrimination. *Inform Theory IRE Trans* 8(2):84–92.
- 15 Balas B, Nakano L, Rosenholtz R (2009) A summary-statistic representation in peripheral vision explains visual crowding. *J Vision* 9(12):13.1–18.
- 16 Freeman J, Simoncelli EP (2011) Metamers of the ventral stream. *Nat Neurosci* 14(9):1195–1201.
- 17 Freeman J, Ziemba CM, Heeger DJ, Simoncelli EP, Movshon JA (2013) A functional and perceptual signature of the second visual area in primates. *Nat Neurosci* 16(7):974–981.
- 18 Yamane Y, Carlson ET, Bowman KC, Wang Z, Connor CE (2008) A neural code for three-dimensional object shape in macaque inferotemporal cortex. *Nat Neurosci* 11(11):1352–1360.
- 19 Arcizet F, Joffrais C, Girard P (2008) Natural textures classification in area V4 of the macaque monkey. *Exp Brain Res* 189(1):109–120.
- 20 David SV, Hayden BY, Gallant JL (2006) Spectral receptive field properties explain shape selectivity in area V4. *J Neurophysiol* 96(6):3492–3505.
- 21 O'Connor DH, Huber D, Svoboda K (2009) Reverse engineering the mouse brain. *Nature* 461(7266):923–929.