# Do computational models of vision need shape-based representations? Evidence from an individual with intriguing visual perceptions

Marcelo Armendariz$^{1,2,3,4,*}$, Will Xiao$^{2,5}$, Kasper Vinken$^{6}$, Gabriel Kreiman$^{1,2}$

[1]Boston Children's Hospital, Harvard Medical School, Boston, MA, USA.
[2]Center for Brains, Minds, and Machines, Cambridge, MA, USA.
[3]Laboratory for Neuro- and Psychophysiology, KU Leuven, Leuven, Belgium.
[4]Leuven Brain Institute, KU Leuven, Leuven, Belgium
[5]Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA
[6]Department of Neurobiology, Harvard Medical School, Boston, MA
*To whom correspondence should be addressed at
Marcelo.ArmendarizGil@childrens.harvard.edu

**Keywords**: ventral visual cortex, intermediate representations, computer vision models, visual deficits

A fundamental task of the human visual cortex is to rapidly and accurately recognize objects in the environment, a task that requires disentangling shape and identity information from incidental features such as size, position, or orientation. A long-standing hypothesis posits that the ventral visual cortex achieves invariant object recognition by transforming the visual input to encode object shape information independently from size, position, etc. It is widely accepted that this transformation happens through a hierarchical sequence of intermediate representations. First, neurons in earlier visual areas detect oriented edges within their receptive fields (Hubel and Wiesel, 1962). This information is combined in downstream regions to form increasingly complex representations from corners in V2, shape and texture in V4 (Pasupathy and Connor, 2002; Kim et al., 2019), and finally, object identity in the inferior temporal cortex (DiCarlo et al., 2012).

Understanding the cascade of transformations underlying perception is a major challenge in visual neuroscience. Findings by Vannuscorps and colleagues (2021) shed new light on this problem by introducing a participant, Davida, who has an abnormal and highly selective impairment in perceiving the orientation of 2D shapes defined by high-contrast edges. From an extensive series of experiments, the authors infer that an intermediate stage of processing (that was presumably affected in the participant) must represent isolated “shapes,” each centered on its own coordinate system defined by the center and major (elongated) and minor axes of the shape. They term this level of representation Intermediate Shape-Centered Representation (ISCR). We focus this commentary on 1) V4 as the proposed neural substrate for ISCRs, and 2) the absence of ISCRs in current computational models of vision.

The authors suggest visual area V4 as the potential locus for ISCRs. Framed within the dual visual stream model of the primate brain (Goodale and Milner, 1992; Mishkin and Ungerleider, 1982), V4 sits at the beginning of the ventral pathway and is densely connected to the dorsal stream (Ungerleider et al., 2008). This watershed position would allow V4 to relay ISCRs to dorsal areas for spatial vision and attention and to inferior temporal areas for visual recognition. V4 receives substantial input from earlier areas (mostly V2) encoding features such as contours and textures, which are useful cues for object segmentation (Anzai et al., 2007) and might facilitate the computation of the proposed ISCRs. While the nature of the transformations that take place in V4 is not yet fully understood, neurons in this region have shown selective tuning to curvatures (Pasupathy and Connor, 1999, 2002; Sharpee et al., 2013), color (Zeki, 1973), texture (Kim et al., 2019), or depth (Livingstone and Hubel, 1988), among others. Previous studies showed that cells in V4 are sensitive to the relative, rather than absolute, position of contour fragments within their receptive fields (Gallant et al., 1996). This observation has led to the proposal that V4 may support the transition from retinotopic coordinates to object-based coordinates (Roe et al., 2012). While relative position representations might be important for any shape-centered scheme, the proposed ISCR is limited to segregated bounded regions of space independent of the relative position of surrounding shapes. For instance, within the ISCR framework, a dashed arrow (composed of slightly unconnected segments) is not perceived as rotated as a whole, but its parts are perceived as rotated instead. In contrast with Davida’s symptoms where object-centered transformations were independent of surrounding segments, neurons in V4 might not exclusively respond to bounded regions alone but also to more complex configurations of unconnected segments that compose a shape (Gallant et al., 1996). Since V4 receives feedback from higher-order regions including parietal and frontal cortex, one possibility is that the scope of local coordinate frames may depend on task demands through top-down modulation. Vannuscorps and colleagues further suggest, based on the differential perception of magno- and parvo-dominant shape attributes, that V4 might comprise "parallel pathways […] specialized in the processing of information derived from the parvocellular and magnocellular channels." However, it remains unclear until what stage along the visual pathway the two channels of information remain segregated. A study in macaque V4 showed that most cells received mixed inputs from both pathways (Ferrera et al., 1994). Finally, another candidate for the neural substrate of ISCRs may be the inferotemporal cortex (IT), the next stage in the ventral pathway after V4. Neurons in the monkey inferotemporal cortex have been reported to code for oriented medial axes in shapes (Hung et al., 2012). Since IT receives most of its inputs from V4, the presumable mixing of parvo- and magno-input in or before V4 remains at odds with the proposed segregated processing.

Independently of the exact locus of the ISCR, a central question is how such representations can be computationally derived from input images. Few studies have examined the role of, or explicitly included, shape-centered representations in computational models of vision. Convolutional neural networks (CNNs), which constitute our current best model of ventral stream processing, do not explicitly use any object-centered representation. Nevertheless, shape-based representations hold promise to address two significant shortfalls of current CNNs. First, current artificial neural networks are vulnerable to so-called adversarial attacks (Szegedy et al., 2014). In adversarial attacks, CNN classification (and internal representation) can be dramatically altered by miniscule changes to the image that are barely noticeable to humans. These changes are distributed throughout the image instead of focused on an object (although adversarial noise can be designed to be localized), suggesting that CNNs fail to weigh shape information appropriately. Since shape is presumably a more robust feature to small image change (Ilyas et al., 2019), emphasizing shape-based representations may improve the robustness of CNNs. Secondly, there is independent evidence that CNNs predominantly use texture instead of shape information when classifying objects, whereas the converse is true for humans (Geirhos et al., 2019). Thus, shape-centered representations may be a useful inductive bias for building future neural networks that are more adversarially robust and more similar to human vision, and should be explored in future work.

Davida's symptoms are intriguing and raise several open questions for further exploration. One unexplained observation is the participant's below-chance performance in several experiments (e.g., figures S2 and S3). In other words, the participant did not so much fail to get the correct answer as she managed to avoid it. This observation indicates that some information about the actual stimulus orientation was available at some point in the visual processing hierarchy. Investigating this perplexing avoidance behavior might provide further insight into Davida's condition. Another intriguing yet underexplored phenomenon is the dynamic aspect of Davida's shape perception. When asked to describe her visual experience when presented with an arrow on a computer screen (movie S1), Davida reported seeing the stimulus rapidly fading in and out of different possible transformations piecemeal, as if multiple representations were competing for perception. Such multistable percepts are, as the authors note, reminiscent of the perceptual effects of binocular rivalry and related phenomena (Blake and Logothetis, 2002). This dynamically changing percept can complicate the interpretation of results because, in most tests, the participant was asked to report only one percept. If the participant indeed perceived multiple rotations and had to choose one to report, it will be interesting to investigate what made the participant choose the actual rotation less frequently than expected by chance. One way to capture the dynamics of Davida's visual experience might be through eye tracking: Does she explore the space occupied by all rotations of the arrow? If so, what percentage of the time, with what transition statistics, and with what relation to the final reported rotation? A third question is whether objects in the entire visual field are affected. Based on results with blurred shapes, it is expected that shapes in the periphery—where visual acuity is low—should be spared. Future experiments like the ones proposed can provide more detailed measurements that help reveal the mechanistic underpinnings of Davida's symptoms.

In conclusion, Vannuscorps and colleagues report on a fascinating neurodevelopmental case that adds support for a shape-centered representation being used as an intermediate step in visual processing in the brain. Although much remains to be learned about the characteristics and mechanisms underlying Davida's symptoms, the authors have gathered a wealth of data that invite interpretation, call for future experiments, and motivate revisiting the role of shape-based features in theories as well as computational models of vision.

## References

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

Blake, R., and Logothetis, N.K. (2002). Visual Competition. Nat. Rev. Neurosci. *3*, 1–11.

DiCarlo, J.J., Zoccolan, D., and Rust, N.C. (2012). How does the brain solve visual object recognition? Neuron *73*, 415–434.

Ferrera, V.P., Nealey, T.A., and Maunsell, J.H.R. (1994). Responses in Macaque Visual Area V4 following Parvocellular and Magnocellular LGN Pathways. J. Neurosci. *14*, 2080–2088.

Gallant, J.L., Connor, C.E., Rakshit, S., Lewis, J.W., and Van Essen, D.C. (1996). Neural responses to polar, hyperbolic, and Cartesian gratings in area V4 of the macaque monkey. J. Neurophysiol. *76*, 2718–2739.

Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F.A., and Brendel, W. (2019). ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. Int. Conf. Learn. Represent. 1–22.

Goodale, M.A., and Milner,  a. D. (1992). Separate visual pathways for perception and action. Trends Neurosci. *15*, 20–25.

Hubel, D.H., and Wiesel, T.N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J. Physiol. *160*, 106–154.

Hung, C.C., Carlson, E.T., and Connor, C.E. (2012). Medial axis shape coding in macaque inferotemporal cortex. Neuron 74, 1099–1113.

Ilyas, A., Santurkar, S., Tsipras, D., Engstrom, L., Tran, B., and Madry, A. (2019). Adversarial Examples are not Bugs , they are Features. Adv. Neural Inf. Process. Syst. *32*.

Kim, X.T., Bair, W., and Pasupathy, A. (2019). Neural Coding for Shape and Texture in Macaque Area V4. J. Neurosci. *39*, 4760–4774.

Livingstone, M.S., and Hubel, D.H. (1988). Segregation of Form, Color, Movement, and Depth: Anatomy, Physiology, and Perception. Science *240*, 740–749.

Mishkin, M., and Ungerleider, L.G. (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. Behav. Brain Res. *6*, 57–77.

Pasupathy, A., and Connor, C.E. (1999). Responses to contour features in macaque area V4. J. Neurophysiol. *82*, 2490–2502.

Pasupathy, A., and Connor, C.E. (2002). Population coding of shape in area V4. Nat. Neurosci. *5*, 1332–1338.

Roe, A.W., Chelazzi, L., Connor, C.E., Conway, B.R., Fujita, I., Gallant, J.L., Lu, H., and Vanduffel, W. (2012). Toward a Unified Theory of Visual Area V4. Neuron *74*, 12–29.

Sharpee, T.O., Kouh, M., and Reynolds, J.H. (2013). Trade-off between curvature tuning and position invariance in visual area V4. Proc. Natl. Acad. Sci. *110*, 11618–11623.

Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., and Fergus, R. (2014). Intriguing properties of neural networks. Int. Conf. Learn. Represent. 1–10.

Ungerleider, L.G., Galkin, T.W., Desimone, R., and Gattass, R. (2008). Cortical Connections of Area V4 in the Macaque. Cereb. Cortex *18*, 477–499.

Vannuscorps, G., Galaburda, A., and Caramazza, A. (2021). Shape-centered representations of bounded regions of space mediate the perception of objects. Cogn. Neuropsychol. DOI: 10.1080/02643294.2021.1960495.

Zeki, S.M. (1973). Colour coding in rhesus monkey prestriate cortex. Brain Res. *53*, 422–427.