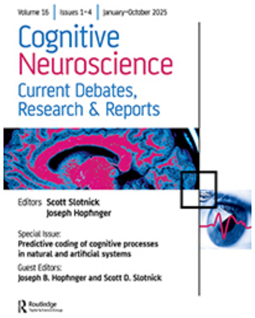

Cognitive Neuroscience
Current Debates, Research & Reports

ISSN: 1758-8928 (Print) 1758-8936 (Online) Journal homepage: www.tandfonline.com/journals/pcns20

Visual input statistics and behavioral relevance jointly constrain higher visual cortex organization

Margaret M. Henderson

COGNITIVE NEUROSCIENCE
https://doi.org/10.1080/17588928.2025.2591254

Routledge
Taylor & Francis Group

COMMENTARY

Visual input statistics and behavioral relevance jointly constrain higher visual cortex organization

Margaret M. Henderson [a,b,c]

[a]Department of Psychology, Carnegie Mellon University, Pittsburgh, PA, USA; [b]Neuroscience Institute, Carnegie Mellon University, Pittsburgh, PA, USA; [c]Machine Learning Department, Carnegie Mellon University, Pittsburgh, PA, USA

ABSTRACT
Ritchie and colleagues propose that the functional organization of higher visual cortex is best understood through the lens of behavioral relevance, advocating for a shift away from theories that center around category selectivity. Building on this, I suggest the statistical structure of visual inputs acts as an additional critical constraint on visual cortex, and that a complete understanding of visual system organization must account for input statistics and how they interact with behavioral relevance. I discuss this using cortical food selectivity as a case study, and additionally describe how deep neural networks can provide new avenues for testing these theories.

KEYWORDS
Vision; object recognition; fMRI; deep neural network; higher visual cortex; natural image statistics

I am in general agreement with the perspective of (Ritchie et al., this issue) namely that the emphasis on category selectivity can be artificially restrictive, and that an ecologically-grounded understanding of visual cortex requires consideration of a wider space of tasks. Many of these tasks require analysis of non-categorical stimulus properties and include dynamic shifts in behavioral relevance (Bracci & de Beeck, 2023; Desimone & Duncan, 1995; Henderson et al., 2025; Hong et al., 2016; Kay et al., 2023).

At the same time, however, visual cortex is also inherently subject to constraints in the 'bottom-up' direction, encompassing the statistical distribution of visual properties in natural inputs, and how those properties co-vary with behavioral relevance. For example, low-level and intermediate visual properties are encoded across higher visual cortex in ways that reflect the co-occurrence statistics of visual and semantic properties: face-selective cortical areas are selective for features that characterize face images, such as low spatial frequency and curved contours (Henderson et al., 2023; Ponce et al., 2017; Srihasam et al., 2014; Yue et al., 2014, 2020), while scene-selective areas are biased toward mid-level properties that characterize scenes, such as cardinal (vertical and horizontal) orientations, rectilinear contours, and high spatial frequency (Henderson et al., 2023; Li & Bonner, 2021; Nasr & Tootell, 2012; Nasr et al., 2014; Rajimehr et al., 2011). In line with Ritchie et al., the interpretation of such findings need not be restricted to a category-selective framework; a more general interpretation is that higher visual regions preferentially represent input components that are relevant for downstream tasks.

In addition to visuo-semantic co-occurrence statistics, input statistics may also constrain higher visual cortex in a more generic way, such that the dimensions encoded across visual cortex capture variance across all inputs, regardless of their semantic meaning or category. For example, cardinal orientation biases in V1 (Appelle, 1972; Barlow, 1961; Girshick et al., 2011; Henderson & Serences, 2021; Li et al., 2003) likely reflect alignment of neural codes with the distribution of features in generic visual inputs. Alignment of neural codes with high-input -variance dimensions may provide a mechanism for efficiently representing stimuli in a format that flexibly supports a range of downstream tasks (Barlow, 1961; Konkle & Alvarez, 2022; Olshausen & Field, 1996). These generic image dimensions may in some cases be correlated with behaviorally-relevant dimensions (e.g., curvature statistics and animacy; Bracci & de Beeck, 2023; Long et al., 2018; Ponce et al., 2017; Yue et al., 2014), but in other cases may be unrelated. Generic input statistics and behavioral relevance thus provide distinct and potentially opposing constraints, with relative importance likely varying between early and higher visual cortex, and between higher areas involved in different task-specific networks.

CONTACT Margaret M. Henderson mmhender@cmu.edu Department of Psychology, Carnegie Mellon University, 5000 Forbes Avenue Pittsburgh, PA, 15213, USA

As a case study for how input statistics and behavioral relevance interact, recent evidence for food-selective regions in ventral visual cortex (Bannert & Bartels, 2022; Jain et al., 2023; Khosla et al., 2022; Pennock et al., 2023) may provide an informative example. The observed food-selective regions overlap partially with color-selective areas (Pennock et al., 2023), potentially reflecting the warm, saturated color statistics associated with food objects (Conway, 2018; Lafer-Sousa et al., 2016; Rosenthal et al., 2018). At the same time, color is not required to elicit food-selective responses (Jain et al., 2023), and, beyond color, there may not be many intermediate visual properties that reliably distinguish food from non-food objects, particularly when considering cross-cultural diversity in food appearance. That is, despite the high behavioral relevance of food, it is unlikely that clustering by similarity in a task-agnostic feature space would be sufficient to give rise to a food-selective dimension in visual cortex. We have argued (Henderson et al., 2025) that nonvisual constraints, including reward processing (Rolls, 2023), multi-modal olfactory and gustatory cues (Avery et al., 2021; Simmons et al., 2005), food-related affordances (Gallivan & Culham, 2015; Mahon & Almeida, 2024), and social cues associated with early learning about food (Amodio & Frith, 2006; Pitcher & Ungerleider, 2021), may serve as additional constraints that interact with visual properties like color, giving rise to food-selective regions in the adult brain. The role of these nonvisual constraints (i.e., behavioral relevance) may play a larger role for food than for other categories, such as faces and words, for which visual similarity may be better aligned with the behavioral relevance of stimuli.

Notably, Ritchie et al. argue that reports of food selectivity may reflect a mis-interpretation, suggesting food-selective regions encode food not as an input category, but in terms of its affordances, citing the similarity between cortical representations of graspable foods and tools (Ritchie et al., 2024) as evidence. In regard to this point, I note that while some foods share affordance properties with some tools (i.e., grasping), the complete space of behaviorally-relevant properties associated with food is non-overlapping with that for tool objects – for example, food is associated with reward circuitry (Rolls, 2023), food is not always directly graspable, and food has distinct taste and smell associations (Avery et al., 2021; Henderson et al., 2025; Simmons et al., 2005). Input statistics also differentiate these categories: food perception may be more dependent on color and material properties relative to tools (Lavin & Hall, 2001; Sato, 2021; Shutts et al., 2009). These non-overlapping constraints suggest partially dissociable cortical circuits supporting food and tool perception, although more empirical work is needed to test this.

More broadly, how can the relative importance of input statistics and behavioral relevance as constraints on visual cortex be disentangled? Recent advances in deep neural networks (DNNs), particularly self-supervised learning (SSL) and multimodal learning, provide new means of testing this. Recent work suggests that training DNNs using SSL on large-scale datasets via methods like contrastive learning (e.g., Chen et al., 2020) can result in feature spaces that align well with higher visual areas (Conwell et al., 2024; Konkle & Alvarez, 2022; Prince et al., 2024; Wang et al., 2023; Zhuang et al., 2021), supporting the theory that task-agnostic input statistics may be sufficient to guide higher visual cortex organization. On the other hand, evidence suggests multimodal language-aligned models (i.e., models that learn to produce similar embeddings for an image and its natural language caption; Radford et al., 2021), better predict fMRI responses in anterior regions of ventral visual cortex compared to self-supervised vision-only models (Wang et al., 2023). This supports a framework in which behaviorally-relevant semantics provide an important constraint on anterior regions. Training multimodal DNNs to jointly model images and captions may provide at least a partial approximation of how the visual system learns to compute behaviorally-relevant properties, as captions include descriptions of rich semantic attributes that are not captured by earlier category-based supervised learning tasks (e.g., ImageNet; Deng et al., 2009). As SSL and multimodal deep learning methods continually advance, these models will provide a rich set of tools for modeling the space of real-world constraints, both input and task-related, that may give rise to biological visual representations.

Acknowledgments

The author thanks Leila Wehbe and Michael Tarr for helpful feedback on this article. Funding support was provided by a startup fund from the Carnegie Mellon Department of Psychology.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by a startup fund from the Carnegie Mellon University Department of Psychology.

ORCID

Margaret M. Henderson http://orcid.org/0000-0001-9375-6680

References

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews Neuroscience*, *7*(*47*), 268–277. https://doi.org/10.1038/nrn1884

Appelle, S. (1972). Perception and discrimination as a function of stimulus orientation: The "oblique effect" in man and animals. *Psychological Bulletin*, *78*(4), 266–278.

Avery, J. A., Liu, A. G., Ingeholm, J. E., Gotts, S. J., & Martin, A. (2021). Viewing images of foods evokes taste quality-specific activity in gustatory insular cortex. Proceedings of the National Academy of Sciences. 118(2): e2010932118.

Bannert, M. M., & Bartels, A. (2022). Visual cortex: Big data analysis uncovers food specificity. *Current Biology*, *32*(19), R1012–R1015. https://doi.org/10.1016/j.cub.2022.08.068

Barlow, H. B. (1961). Possible principles underlying the transformations of sensory messages. In W. A. Rosenblith (Ed.), *Sensory communication* (Vol. 13, pp. 217–234). Chap. MIT Press.

Bracci, S., & de Beeck, H. P. O. (2023). Understanding human object vision: A picture is worth a thousand representations. *Annual Review of Psychology*, *74*(1), 113–135. https://doi.org/10.1146/annurev-psych-032720-041031

Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A Simple Framework for Contrastive Learning of Visual Representations. International Conference on Machine Learning Vienna, Austria. https://arxiv.org/abs/2002.05709

Conway, B. R. (2018). The organization and operation of inferior temporal cortex. *Annual Review of Vision Science*, *4*(1), 381–402. https://doi.org/10.1146/annurev-vision-091517-034202

Conwell, C., Prince, J. S., Kay, K. N., Alvarez, G. A., & Konkle, T. (2024). A large-scale examination of inductive biases shaping high-level visual representation in brains and machines. *Nature Communications*, *15*(9383). https://doi.org/10.1038/s41467-024-53147-y

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. IEEE Conference on Computer Vision and Pattern Recognition. Miami, FL, 248–255.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222. https://doi.org/10.1146/annurev.ne.18.030195.001205

Gallivan, J. P., & Culham, J. C. (2015). Neural coding within human brain areas involved in actions. *Current Opinion in Neurobiology*, *33*, 141–149. https://doi.org/10.1016/j.conb.2015.03.012

Girshick, A. R., Landy, M. S., & Simoncelli, E. P. (2011). Cardinal rules: Visual orientation perception reflects knowledge of environmental statistics. *Nature Neuroscience*, *14.7*(7), 926–932. https://doi.org/10.1038/nn.2831

Henderson, M. M., Tarr, M. J., & Wehbe, L. (2025). Origins of food selectivity in human visual cortex. *Trends in Neurosciences*, *48*(2), 113–123. https://doi.org/10.1016/j.tins.2024.12.001

Henderson, M. M., & Serences, J. T. (2021). Biased orientation representations can be explained by experience with non-uniform training set statistics. *Journal of Vision*, *21*(8), 10–10. https://doi.org/10.1167/jov.21.8.10

Henderson, M. M., Serences, J. T., & Rungratsameetaweemana, N. (2025). Dynamic categorization rules alter representations in human visual cortex. *Nature Communications*, *16*(1), 3459. https://doi.org/10.1038/s41467-025-58707-4

Henderson, M. M., Tarr, M. J., & Wehbe, L. (2023). Low-level tuning biases in higher visual cortex reflect the semantic informativeness of visual features. *Journal of Vision*, *23*(4), 8–8. https://doi.org/10.1167/jov.23.4.8

Hong, H., Yamins, D. L. K., Majaj, N. J., & DiCarlo, J. J. (2016). Explicit information for category-orthogonal object properties increases along the ventral stream. *Nature Neuroscience*, *19*(4), 613–622. https://doi.org/10.1038/nn.4247

Jain, N., Wang, A., Henderson, M. M., Lin, R., Prince, J. S., Tarr, M. J., & Wehbe, L. (2023). Selectivity for food in human ventral visual cortex. *Communications Biology*, *6*(125).

Kay, K., Bonnen, K., Denison, R. N., Arcaro, M. J., & Barack, D. L. (2023). Tasks and their role in visual neuroscience. *Neuron*, *111*(11), 1697–1713. https://doi.org/10.1016/j.neuron.2023.03.022

Khosla, M., Murty, N. A. R., & Kanwisher, N. (2022). A highly selective response to food in human visual cortex revealed by hypothesis-free voxel decomposition. *Current Biology*, *32*(19), 4159–4171.e9. https://doi.org/10.1016/j.cub.2022.08.009

Konkle, T., & Alvarez, G. A. (2022). A self-supervised domain-general learning framework for human ventral stream representation. *Nature Communications 13*(491).

Lafer-Sousa, R., Conway, B. R., & Kanwisher, N. G. (2016). Color-biased regions of the ventral visual pathway lie between face- and place-selective regions in humans, as in macaques. *The Journal of Neuroscience*, *36*(5), 1682–1697. https://doi.org/10.1523/JNEUROSCI.3164-15.2016

Lavin, T. A., & Hall, D. G. (2001). Domain effects in lexical development: Learning words for foods and toys. *Cognitive Development*, *16*(4), 929–950. https://doi.org/10.1016/S0885-2014(02)00070-9

Li, B., Peterson, M. R., & Freeman, R. D. (2003). Oblique effect: A neural basis in the visual cortex. *Journal of Neurophysiology*, *90*(1), 204–217. https://doi.org/10.1152/jn.00954.2002

Li, D. S. P., & Bonner, M. F. (2021). Emergent selectivity for scenes, object properties, and contour statistics in feedforward models of scene-preferring cortex. *BioRxiv*. https://doi.org/10.1101/2021.09.24.461733

Long, B., Yu, C. P., & Konkle, T. (2018). In *Proceedings of the National Academy of Sciences of the United States of America* (Vol. 115, pp. E9015–E9024).

Mahon, B. Z., & Almeida, J. (2024). Reciprocal interactions among parietal and occipitotemporal representations support everyday object-directed actions. *Neuropsychologia*, *198*, 108841. https://doi.org/10.1016/j.neuropsychologia.2024.108841

Nasr, S., Echavarria, C. E., & Tootell, R. B. H. (2014). Thinking outside the box: Rectilinear shapes selectively activate scene-selective cortex. *The Journal of Neuroscience*, *34*(20), 6721–6735.

Nasr, S., & Tootell, R. B. H. (2012). "A cardinal orientation bias in scene-selective visual cortex". *The Journal of Neuroscience*, *32*(43), 14921–14926. https://doi.org/10.1523/JNEUROSCI.2036-12.2012

Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*(6583), 607–609. https://doi.org/10.1038/381607a0

Pennock, I. M. L., Racey, C., Allen, E. J., Wu, Y., Naselaris, T., Kay, K. N., Franklin, A., & Bosten, J. M. (2023). Color-biased regions in the ventral visual pathway are food selective. *Current Biology: CB*, *33*(1), 134–146.e4. https://doi.org/10.1016/j.cub.2022.11.063

Pitcher, D., & Ungerleider, L. G. (2021). Evidence for a third visual pathway specialized for social perception. *Trends in Cognitive Sciences*, *25*(2), 100–110. https://doi.org/10.1016/j.tics.2020.11.006

Ponce, C. R., Hartmann, T. S., & Livingstone, M. S. (2017). End-stopping predicts curvature tuning along the ventral stream. *The Journal of Neuroscience*, *37*(3), 648–659.

Prince, J. S., Alvarez, G. A., & Konkle, T. (2024). Contrastive learning explains the emergence and function of visual category-selective regions. *Science Advances*, *10*(39), 10.39. https://doi.org/10.1126/sciadv.adl1776

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning Transferable Visual Models From Natural Language Supervision. 38th International Conference on Machine Learning (Vol. 139). https://doi.org/10.48550/arXiv.2103.00020

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 Biology*, *9*(4), 1000608. https://doi.org/10.1371/journal.pbio.1000608

Ritchie, J. B., Andrews, S. T., Vaziri-Pashkam, M., & Baker, C. I. (2024). Graspable foods and tools elicit similar responses in visual cortex. *Cerebral Cortex*, *34*(9), bhae383. https://doi.org/10.1093/cercor/bhae383

Rolls, E. T. (2023). The orbitofrontal cortex, food reward, body weight and obesity. *Social Cognitive and Affective Neuroscience*, *18*(1), nsab044. https://doi.org/10.1093/scan/nsab044

Rosenthal, I., Ratnasingam, S., Haile, T., Eastman, S., Fuller-Deets, J., & Conway, B. R. (2018). Color statistics of objects, and color tuning of object cortex in macaque monkey. *Journal of Vision*, *18*(11), 1–1. https://doi.org/10.1167/18.11.1

Sato, W. (2021). Color's indispensable role in the rapid detection of food. *Frontiers in Psychology*, *12*, 5442. https://doi.org/10.3389/fpsyg.2021.753654

Shutts, K., Condry, K. F., Santos, L. R., & Spelke, E. S. (2009). Core knowledge and its limits: The domain of food. *Cognition*, *112*(1), 120–140. https://doi.org/10.1016/j.cognition.2009.03.005

Simmons, W. K., Martin, A., & Barsalou, L. W. (2005). Pictures of appetizing foods activate gustatory cortices for taste and reward. *Cerebral Cortex*, *15*(10), 1602–1608. https://doi.org/10.1093/cercor/bhi038

Srihasam, K., Vincent, J. L., & Livingstone, M. S. (2014). Novel domain formation reveals proto-architecture in inferotemporal cortex. *Nature Neuroscience*, *17*(12), 1776–1783. https://doi.org/10.1038/nn.3855

Wang, A. Y., Kay, K., Naselaris, T., Tarr, M. J., & Wehbe, L. (2023). Better models of human high-level visual cortex emerge from natural language supervision with a large and diverse dataset. *Nature Machine Intelligence*, *5*(*12* 5), 1415–1426. https://doi.org/10.1038/s42256-023-00753-y

Yue, X., Pourladian, I. S., Tootell, R. B. H., & Ungerleider, L. G. (2014). Curvature-processing network in macaque visual cortex. In *Proceedings of the National Academy of Sciences of the United States of America* (Vol. 111, p. E3467).

Yue, X., Robert, S., & Ungerleider, L. G. (2020). Curvature processing in human visual cortical areas. *Neuroimage*, *222*, 117295. https://doi.org/10.1016/j.neuroimage.2020.117295

Zhuang, C., Yan, S., Nayebi, A., Schrimpf, M., Frank, M. C., DiCarlo, J. J., & Yamins, D. L. K. (2021). Unsupervised neural network models of the ventral visual stream. In *Proceedings of the National Academy of Sciences of the United States of America* (Vol. 118).