Do retinal neurons also represent somatosensory inputs? On why neuronal responses are not sufficient to determine what neurons do

Lotem Elber-Dorozko[1,2] and Yonatan Loewenstein[3-6]

1. Department of Humanities and Arts, Technion, Haifa, Israel

2. The Center for Philosophy of Science, University of Pittsburgh, Pittsburgh, PA, USA

3. The Edmond and Lily Safra Center for Brain Sciences, The Hebrew University, Jerusalem, Israel

4. Department of Neurobiology, The Alexander Silberman Institute of Life Sciences, The Hebrew University, Jerusalem, Israel

5. Department of Cognitive Sciences, The Hebrew University, Jerusalem, Israel

6. The Federmann Center for the Study of Rationality, The Hebrew University, Jerusalem, Israel

Abstract

How does neuronal activity give rise to cognitive capacities? To address this question, neuroscientists hypothesize about what neurons 'represent', 'encode', or 'compute', and test these hypotheses empirically. This process is similar to the assessment of hypotheses in other fields of science and as such is subject to the same limitations and difficulties that have been discussed at length by philosophers of science. In this paper, we highlight an additional difficulty in the process of empirical assessment of hypotheses that is unique to the cognitive sciences. We argue that, unlike in other scientific fields, comparing hypotheses according to the extent to which they explain or predict empirical data can lead to absurd results. Other considerations, which are perhaps more subjective, must be taken into account. We focus on one such consideration, which is the purposeful function of the neurons as part of a biological system. We believe that progress in neuroscience critically depends on properly addressing this difficulty.

Paper

Much of neuroscientific research is focused on identifying what neurons represent, even if the definition of representation is debated. In practice, if neuronal activity correlates or causally relates to the relevant variable, the hypothesis about 'representation'[1] of that variable is taken to be

[1] For brevity we discuss representation alone, but similar points can be made regarding computation

supported. For example, we believe that the somatosensory cortex encodes information about somatosensory stimulation. Some observations that support this are that somatosensory stimulation elicits response in somatosensory cortical regions (e.g., Saadon-Grosman, Loewenstein, and Arzy 2020) and that electrical stimulation of these cortical regions elicits a report of feeling somatosensory stimulation (e.g., Penfield and Boldrey 1937). Similar experiments underlie the beliefs that faces are represented in the fusiform gyrus (Kanwisher, McDermott, and Chun 1997), that 'reward prediction error' is encoded in the ventral tegmental area (Schultz, Dayan, and Montague 1997), and many similar beliefs.

Neuroscientists may disagree about neural representations.[2] Consider an experiment in which a neuron has been shown to respond when images of faces are presented to a participant. One neuroscientist may argue that a face-representing neuron was identified. Nevertheless, there are alternative interpretations. A second neuroscientist may argue that the neuron represents oval shapes. A third neuroscientist may argue that the neuron responded only because a nose has been presented. A common way of testing these competing hypotheses is to conduct additional experiments, in which the competing hypotheses predict different patterns of neuronal activity. For example, present non-face oval objects, non-oval faces, and noses in isolation to the participant.

However, this approach can lead to absurd conclusions. To see why, consider the question of what retinal ganglion cells represent. It is generally accepted that these cells represent a spot of light at a particular spatial angle (for simplicity we ignore center-surround properties). It is well-known that pressing the eyeball elicits light sensations also known as 'phosphenes' (Grüsser and Hagner 1990). Taking these phosphenes as evidence for the activation of retinal ganglion cells in response to somatosensory stimulation, one may hypothesize that retinal ganglion cells represents both visual and somatosensory aspects of the world. If we were to test the competing hypotheses by recording the activity of retinal ganglion cells in response to somatosensory stimuli, we would be compelled to judge in favor of the visual+somatosensory representation hypothesis. Nonetheless, neuroscientists take the visual representation hypothesis to be correct and the visual+somatosensory hypothesis to be absurd. Similarly, while the activity of the vestibular

[2] We assume that neuronal representations are independent of scientific interests and explanations. For alternative views (Cao 2022; Hacohen 2022; Egan 2014)

system is affected by alcohol ingestion (Tianwu et al. 1995), it is not believed that the vestibular system also *represents* the recent history of alcohol consumption.

It has been argued before that a neuron's response to a stimulus is not sufficient to claim that it *represents* it (Dretske 1988, chap. 3; Cao 2022; Baker, Lansdell, and Kording 2021). Moreover, even an analysis of a multi-area pattern of activation is not sufficient to determine between two different hypotheses about representation (Gessell, Geib, and De Brigard 2021). Here we make an even stronger claim: no amount of neuronal data, by itself, can be used to conclusively decide between two competing hypotheses about representation. For the hypothesis that is more strongly supported by the data, such as the visual+somatosensory representation hypothesis, can be taken to be unlikely for other reasons.

We claim that, in practice, different considerations guide us when comparing hypotheses about neuronal representations. One reason why the neuroscience community takes retinal ganglion cells to represent visual input but not somatosensory input is because it is believed that the eye has been adapted for visual processing. More generally, much neuroscientific work is guided by the intuition that representations are necessarily linked with purposeful, teleological *function*. According to popular philosophical views, representations (Neander 2017; Millikan 1989; Dretske 1988) and computations (Piccinini 2015; Fresco 2021) are determined according to the functions they perform,[3] and some functions make more sense than others. A well-known example in philosophy is that the heart's function is to pump blood and not to make thumping sounds (Wouters 2005). Similarly, we can say that retinal ganglion cells do not represent somatosensory input because they do not have the function of relaying somatosensory information.

The difficulty with biological function is that there is no clear-cut way to know the function of a system or sub-system. It is often posited that biological functions are relevant to survival and reproduction (Millikan, 1989; Piccinini, 2015, Chapter 6; but see Cummins, 1975). While these guidelines allow us to exclude certain hypothesized functions, such as making thumping sounds, they still leave much room for various possible functions assigned to the same system.

In the examples of retinal ganglion cells and the vestibular system, neuroscientific intuition is relatively clear. However, there are many cases where neuroscientists may disagree regarding

[3] In contrast to these references, we aim to describe how representations can be identified (epistemology) and make no claims about what they are exactly (ontology).

function. Consider a neuron in the visual cortex that responds exclusively to elongated, rope-like shapes. One neuroscientist may hypothesize that the role of some neurons in the visual cortex is to facilitate avoidance of dangerous stimuli. On this view, the recorded neuron may represent snakes. The fact that a similar object, say a garden hose, elicits a similar response is interpreted as a misrepresentation, just like the vestibular system's response to alcohol is a misrepresentation. Another scientist may assume that neurons in the visual cortex aim to accurately represent the visual scene, of which elongated objects are occasionally a part. On this view, there is no misrepresentation. Crucially, if one is convinced by the previous examples that response properties may differ from representation, it becomes evident that the question of whether this neuron represents 'snakes' or 'elongated objects' cannot be resolved solely by conducting more experiments. While response properties play an important part in identifying what neurons represent, alone they are not decisive, as each neuroscientist may legitimately argue that some response properties are 'misrepresentations'. What will settle this debate is agreement on the function of the neuron.

One invaluable way to understand neuronal activity is to examine it in the context of behavior (Krakauer et al. 2017; Niv 2021). Nonetheless, behavioral response cannot replace nor determine 'function'. To see the usefulness, but also the limitations of behavioral experiments, consider a sea-creature with receptors that, like retinal ganglion cells, respond both to light and touch. Studying the creature's behavior, we discover that it moves in the direction of the activated receptors, regardless of the source of activation. As both visual and somatosensory inputs lead to a behavioral response, the behavioral experiment is consistent with various hypotheses about representation. Therefore, behavior per-se is not sufficient to determine representation. However, if we discover that in the course of evolution, the food of the creature's ancestors tended to be closer to well-lit areas, we may conclude that its phototactic behavior has a function. If no analogous function for somatosensory-guided movement is found, we may say that the creature's receptors are light receptors, while their response to touch is a 'side-effect' of a mechanism whose purpose is to move the creature to well-lit areas. This insight thus required us to study both the creature's behavior and to speculate about the effect of this behavior throughout evolution [(Krakauer et al. 2017) argue similarly].

The argument presented here indicates that neuroscientific data alone may not suffice to resolve scientific disagreements about representation. How can such disagreements be resolved? We suggest that an explicit discussion of the considerations of function that play a part in the various stages of neuroscientific practice will help clarify disagreements. Such discussions can benefit from the philosophical literature that distinguishes between different descriptions of biological function, whether it is determined by natural-selection, other historical properties, or current contribution to goals (Wouters 2005). Adopting a clear stance on biological function can also help come up with ways to address questions about function empirically, e.g., by examining evolutionary changes (Cisek and Hayden 2022), or how environmental change affects behavior on different time scales. Moreover, serious consideration of biological function is important to avoid telling just-so stories, which describe functions as evolutionarily selected without justification (Gould and Lewontin 1979). We are hopeful that shining a spotlight on biological function as a central feature guiding scientific decisions will help create more precise ways of thinking about representation in the brain.

Acknowledgements

We thank Yoram Burak, Ori Hacohen, Arnon Levy, Shimon Marom, Israel Nelken, Orly Shenker, executive editor Rick Dale and two anonymous editors at 'Cognitive Science' for their careful reading of the manuscript and their insightful comments. This work was supported by the Gatsby Charitable Foundation and the DFG (CRC 1080).

References

Baker, Ben, Benjamin Lansdell, and Konrad Kording. 2021. "A Philosophical Understanding of Representation for Neuroscience." *Arxiv*. https://doi.org/arXiv:2102.06592v2.

Cao, Rosa. 2022. "Putting Representations to Use." *Synthese* 200 (2): 151. https://doi.org/10.1007/s11229-022-03522-3.

Cisek, Paul, and Benjamin Y Hayden. 2022. "Neuroscience Needs Evolution." *Philosophical Transactions of the Royal Society B: Biological Sciences* 377 (1844): 20200518. https://doi.org/10.1098/rstb.2020.0518.

Cummins, Robert. 1975. "Functional Analysis." *The Journal of Philosophy* 72: 741–65.

Dretske, Fred. 1988. *Explaining Behavior: Reasons in a World of Causes.* The MIT Press.

Egan, Frances. 2014. "How to Think about Mental Content." *Philosophical Studies* 170 (1): 115–35. https://doi.org/10.1007/s11098-013-0172-0.

Fresco, Nir. 2021. "Long-Arm Functional Individuation of Computation." *Synthese* 199 (5): 13993–16. https://doi.org/10.1007/s11229-021-03407-x.

Gessell, Bryce, Benjamin Geib, and Felipe De Brigard. 2021. "Multivariate Pattern Analysis and the Search for Neural Representations." *Synthese* 199 (5): 12869–89. https://doi.org/10.1007/s11229-021-03358-3.

Gould, S. J., and R. C. Lewontin. 1979. "The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme." *Proceedings of the Royal Society of London.* 205: 581–98.

Grüsser, Otto-Joachim, and Michael Hagner. 1990. "On the History of Deformation Phosphenes and the Idea of Internal Light Generated in the Eye for the Purpose of Vision." *Documenta Ophthalmologica* 74 (1): 57–85. https://doi.org/10.1007/BF00165665.

Hacohen, Ori. 2022. "What Are Neural Representations? A Cummins Functions Approach." *Philosophy of Science*, 1–20. https://doi.org/DOI: 10.1017/psa.2022.6.

Kanwisher, Nancy, Josh McDermott, and Marvin M Chun. 1997. "The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception." *The Journal of Neuroscience* 17 (11): 4302 LP – 4311. https://doi.org/10.1523/JNEUROSCI.17-11-

04302.1997.

Krakauer, John W, Asif A Ghazanfar, Alex Gomez-Marin, Malcolm A MacIver, and David Poeppel. 2017. "Neuroscience Needs Behavior: Correcting a Reductionist Bias." *Neuron* 93 (3): 480–90. https://doi.org/10.1016/j.neuron.2016.12.041.

Millikan, Ruth Garrett. 1989. "In Defense of Proper Functions." *Philosophy of Science* 56 (2): 288–302. http://www.jstor.org/stable/187875.

Neander, Karen. 2017. *A Mark of the Mental*. The MIT Press.

Niv, Yael. 2021. "The Primacy of Behavioral Research for Understanding the Brain." *Behavioral Neuroscience* 135 (5): 601–9. https://doi.org/10.1037/bne0000471.

Penfield, Wilder, and Edwin Boldrey. 1937. "Somatic Motor and Sensory Representation in the Cerebral Cortex of Man As Studied by Electrical Stimulation." *Brain* 60 (4): 389–443. https://doi.org/10.1093/brain/60.4.389.

Piccinini, Gualtiero. 2015. *Physical Computation: A Mechanistic Account*. Oxford University Press.

Saadon-Grosman, Noam, Yonatan Loewenstein, and Shahar Arzy. 2020. "The 'Creatures' of the Human Cortical Somatosensory System." *Brain Communications* 2 (1): fcaa003. https://doi.org/10.1093/braincomms/fcaa003.

Schultz, W., P. Dayan, and P. R. Montague. 1997. "A Neural Substrate of Prediction and Reward." *Science* 275: 1593–99. https://doi.org/10.1126/science.275.5306.1593.

Tianwu, H, Y Watanabe, M Asai, K Shimizu, S Takada, and K Mizukoshi. 1995. "Effects of Alcohol Ingestion on Vestibular Function in Postural Control." *Acta Oto-Laryngologica.* 519: 127–31. https://doi.org/10.3109/00016489509121886.

Wouters, Arno. 2005. "The Function Debate In Philosophy." *Acta Biotheoretica* 53: 123–51.