What can be inferred from a ring in the fly brain? A case for historical functions in neuroscience

Lotem Elber-Dorozko

Abstract

It is generally accepted that neural properties play an important role in supporting cognitive models. But it is less clear how this is done. Several authors suggested that neural evidence supports cognitive models through inference to the best explanation. This paper argues that for such inference to work, scientists should adopt a historical approach to functions because neural properties can only be explained by their causal history. The method by which such inferences can take place is exemplified on a case study of the ring-attractor model for head-direction representation, and the neural evidence for it in flies and in mice.

1. Introduction

One central aim of cognitive science is to explain how cognitive phenomena, such as object recognition, navigation, long-term memory, and more, are performed. Generally, such explanations appeal to the mechanisms underlying those capacities, understood broadly to include complex dynamics, population activity, and network interactions (Boone and Piccinini 2016; Craver 2007). Despite philosophical arguments for the autonomy of psychology from neuroscience (Fodor 1968, 1974; Shapiro 2017), it is generally agreed today, at least among scientists, that neuroscientific evidence is crucial, if not necessary, in supporting hypotheses about how cognitive capacities are performed. This stance is supported by the argument that a suggested model is merely a possible explanation until it has been demonstrated to be physically implemented in the target system (Piccinini and Craver 2011; Cummins 2000; Craver 2007).

But how can neuroscientific evidence be used to evaluate suggested models? Several authors suggested that the best way to conceive of the support neural evidence provides is through justificatory abduction, also known as inference to the best explanation (henceforth IBE; Calzavarini and Cevolani 2022; Coraci et al. 2024). This paper argues that, on this premise, the type of cognitive functions supported by neural evidence must be understood etiologically, i.e., grounded in the causal history of the system, rather than as current causal roles. This argument is exemplified on a comparative case study – the ring-attractor model is an abstract dynamical model that aims to explain the cognitive capacity to represent head-direction. Neural correlates of the model have been identified both in Drosophila flies and in mice. The evidence in the former case is intuitively much more convincing, and this paper argues that this result naturally emerges from the account presented here. The upshot is that neuroscientific practice should shift its focus from neural activity towards organizational and developmental properties of neural structures.

The second section describes IBE in the context of neuroscience; the third argues that IBE in neuroscience requires a historical view of function; the fourth section analyzes the ring-attractor model as a case study; the fifth further supports IBE as the right framework in cognitive neuroscience; the sixth section briefly concludes.

2. IBE in neuroscience

There are several factors that hinder scientific efforts to relate neural structures and cognitive functions. First, it is common to identify many-to-many relationships between functional and structural properties when examining the brain (McCaffrey 2023; Anderson 2010; Burnston 2016; Viola 2021). Additionally, practically always, it is still unknown what is the right kind of neural property to be examined as potential physical implementations, whether they are single spike forms, firing rate, population activity, trajectories in some dynamical phase space, or else (Maley 2018; Anderson 2010). Finally, there is also debate about cognitive ontology: whether cognition can be described as a set of distinct cognitive capacities, and if so, which capacities those are (McCaffrey and Wright 2022).

Despite these challenges, it is quite common for scientists to identify a correlation between neural activity and some property or task, and to take these results as evidence that the relevant neural properties are involved in the relevant task, usually by representing the associated property. Given said challenges in relating structure and function, the legitimacy of such inferences has been put under pressure by both scientists and philosophers (Marom et al. 2009; Jonas and Kording 2017; Guest and

Martin 2023; Poldrack 2006). Granted, support from neural evidence to cognitive models is not deductive, nor is it simple few-to-many inductive inference. Furthermore, as I argue in section five, there are good reasons to believe that it does not incorporate other popular forms of confirmation. Thus, a concrete account of the way neural evidence can support (or decrease support for) cognitive models is in order.

Filling this lacuna, several authors suggested that the inference of a brain area's function from data about its activity is an inference to the best explanation (Aizawa 2026; Coraci et al. 2024; Calzavarini and Cevolani 2022). This inference can take place in two ways: first, neural activity may be associated with some function because this association best explains *the cognitive capacity*, and second, neural activity may be associated with some function because this association best explains *the neural activity*. The latter, but not the former, is a case of support for a cognitive model from neural evidence.

Aizawa (2026) suggests that scientists conduct the first kind of inference. Regarding place cells in the hippocampus, he writes: "the reason to believe that the hippocampus, rather than some other brain structure, is the cognitive map is that the place cells would explain how the hippocampus could be a cognitive map" (p. 9). Continuing this line of thought, one can add that a reason to believe that some hippocampal cells represent place rather than a different variable is because the representation of place best explains the capacity to navigate by use of a cognitive map. Such inferences surely take place in neuroscientific practice, but they are not cases of support from neural evidence to cognitive models. In IBE, the explananda support the explaining hypothesis. Applied in this case, the explained capacity to navigate by use of a cognitive map supports the hypothesis that neurons represent place. Thus, in this form of IBE, the cognitive capacity supports claims about neural activity. For this reason, the focus of this paper is on the second type of IBE in neuroscience.

The second type of inference to the best explanation pertains to explanation of the scientific results. Somewhat simplified, the idea is that what best explains the result that neural activity is correlated with some variable, is the fact that the function of neural activity is to correlate with said variable. Such inferences are often transparent, and scientists do not try to justify them; neurons whose activity is modulated by orientation are 'orientation neurons', i.e., neurons whose function is to carry information about orientation, and neurons that respond to faces more than to other objects are 'face neurons', i.e., neurons whose function is to respond to faces, etc. Such inferences can naturally be described as inferences to the best explanation; upon discovering that neurons have some response properties, the best explanation seems to be that they are meant to have this response property.

Such inferences are clearly defeasible and their power depends on the details. Specifically for fMRI (Functional Magnetic Resonance Imaging) studies, such inferences have been criticized. Poldrack (2006) calls this reverse inference and points out that it is not deductively valid (as all cases of Inference to the Best Explanation are). Coraci and colleagues (2024) suggest that such inferences should not be an all-or-nothing matter, and that the inferred function can have different degrees of support depending on various properties of the evidence and the cognitive model.

3. Inferring the right kind of functions

Much of this debate hinges on a question that has not received enough attention in the literature; what is meant by claims that neural properties have functions? If neural function is inferred because it best explains neural activity, then an account of such functions should be suggested, where the manner in which they explain neural activity is understood. While scientists are usually not explicit about their view of 'functions', their position is often evident in their writings. Several scientists, such as (Poldrack 2006), seem to adopt a systemic view of biological function. This view, championed by Cummins (1975) and later Craver (2013), is accepted by many today. On this view, a property has a function Φ if it can behave in manner Φ, and this behavior is part of an analytic account of some capacity. That is, having a function is simply playing some causal role in a more general capacity of the system that scientists aim to explain. In line with this view, Poldrack (2006) describes reverse inference as inference that "the activity of area Z in the present study demonstrates engagement of cognitive process X by task comparison A." (p. 59). Thus, Poldrack and others, view the function of a brain area as its engagement in some cognitive process.

However, there is a problem with the use of this account of function within the context of explaining neural activity: this type of function simply cannot explain why neural activity is the way it is. First, note that under Cummins' (1975) original description of biological function as systemic, this function plays a role in the explanation of the general capacity. Functions are part of an analytic account that explains how some biological capacity takes place. They are not meant to explain why systems and their parts have certain properties. In fact, Cummins is highly critical of attempts to explain properties by appeal to function, which is the reason he separates his account of function from natural selection.

Additionally, simply considering the form these explanations would take, it seems evident that they are not up for the task. As an example, consider an explanation of why the activity of neurons in early visual cortex is modulated by the orientation of visual stimuli. On this account of function, their activity is modulated by orientation because they play a causal role in object recognition. But this explanation clearly confuses explanans and explananda. It is like saying that Mercury expands when heated because it is used in thermometers, or that diamonds are beautiful because they are used in jewelry. In both of these cases, the objects can play a role *because* of their properties, not vice versa. Similarly for neural activity, these neurons can play a part in object recognition because they are modulated by properties of visual input such as orientation, and not vice versa.

One may suggest that the upshot of this analysis is that neuroscientists should give up attempts to explain neural activity, and simply appeal to it to strengthen or weaken support in their theories. This is not an outrageous suggestion outright. After all, scientific interest in neural activity is mainly motivated by scientific desire to understand cognition and cognitive capacities. Most studies about neural activity are designed and interpreted in this context. Nonetheless, upon further consideration, much is lost if scientists give up the attempt to explain neural activity and neural properties. Let us put aside questions about the degree of interest scientists should have in neural properties regardless of cognitive capacities. Recall how IBE was first suggested. The question it aimed to answer is - how can evidence of neural activity increase or diminish support of a suggested model of a cognitive capacity. If one gives up the idea that scientists should aim to explain neural properties, then one should come up with an alternative framework for the relationship between neural properties and cognitive models, so that the former can serve as supporting evidence for the latter (see also section five).

Before giving up on IBE, I suggest that a smaller amendment can be made to this framework and that is a change in how purposeful functions are conceptualized. First, note that today it is well accepted that the vast majority of explanations in science are causal (Woodward 2003). Therefore, it is only natural to attempt to explain neural properties by appeal to their causal history. Many philosophers and scientists prefer to avoid explicit appeals to the history of the system when describing what it does, often with the argument that such history is not directly available. But there is a leading philosophical framework for function that appeals to the history of selection of properties, which is the etiological framework (Neander 1991; Garson 2017; Millikan 1989). On this framework, somewhat simplistically, the function of a property or object is the aspect that historically led to its selection or the selection of its containing organism. This view of function squares nicely with IBE because IBE suggests how one can draw conclusions about the history of the system without having direct evidence about it: the history is inferred as the best explanation of current neural properties.

This paper does not have the scope to go into the question of what specific history of selection a function must have (a question thoroughly debated in the literature: Neander 1991; Garson 2017; Millikan 1989). Regardless of the specific requirements, a historical approach to function allows the explanation of neural properties. Neural properties are explained as the result of etiological processes: they exist today either because they were selected or for other evolutionary reasons, such as interaction with selected properties, structural constraints or neutral drift (Novick 2023).

4. The ring-model for head-direction representation in flies and mice

The manner in which causal history can be inferred from neural properties can best be understood through an example. A central component of navigation is representation of head-direction, i.e., which way the animal is facing in a 360° circle. The question of how animals represent head-direction has received much attention.

Experimentalists discovered a subset of neurons in the rat postsubiculum that is modulated by head-direction (Taube et al. 1990). Inspired by such findings, theoreticians developed a dynamical model of head-direction representation, in which head-direction is represented as location on a ring attractor (Zhang 1996). This means that in an abstract space defined by the dynamical equations, there is a ring (a continuous closed line) of stable states. Each state corresponds to a specific head-direction, so that nearby states on the dynamical ring correspond to nearby head-directions. Without any inputs the model will remain in the same state. Mixed states are not stable, so one cannot be in a stable state representing two head-directions at once. The state can be changed according to external inputs, either continuously with inputs such as speed, or discontinuously with inputs such as visual cues. Thus, this model fits nicely with the requirements for a system for head-direction representation: only one direction can be represented at a time, direction remains fixed with no inputs, and various inputs can change the representation as needed.

Is this ring-attractor model a good description of how head-direction is represented in brains? Experimentalists use findings about neural properties to answer this question. In the Drosophila fly, they discovered the striking result that the different abstract stable states of the ring-attractor model can be associated with activation of neurons in an anatomical structure that is an actual physical ring, termed the ellipsoid body (Fig.

1, left), so that there is a strong correlation between the fly's head-direction and neuronal activity in a specific location in this physical ring. Furthermore, activity follows the ring-attractor model: there is a stable single ‘bump’ of activity on the ring, so that only neurons at one location are active at a time (Turner-Evans et al. 2020; Kim et al. 2017). To reiterate, in the brain of the Drosophila fly, there is an organ organized as an actual physical ring, in which the activity of neurons in different locations on the ring corresponds to the head-direction of the fly (Fig. 1).

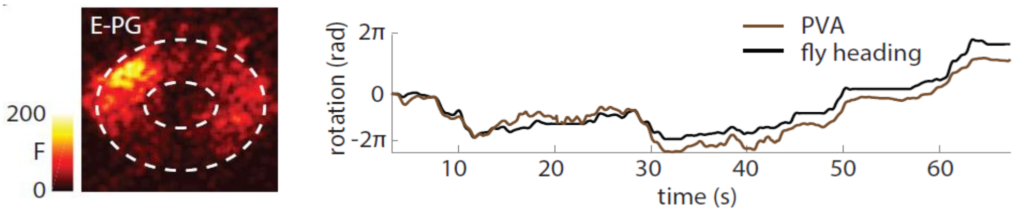

Figure 1. Adapted from (Turner-Evans et al. 2017, Fig. 1) under a CC-BY license. Correlates of the ring model in the fly ellipsoid body. Left: two-photon calcium imaging of the ellipsoid body. Activity in this ring has the structure of a single bump - strong activity in a specific location on the ring and weak activity elsewhere. Right: example of the correlation between neuronal activity and fly heading (i.e., head-direction). PVA – population vector average, i.e., the angle computed by averaging activity strength over the entire ring, so that each slice of the ring denotes a specific angle.

This finding was received with much excitement. Even those less familiar with neuroscientific research are captivated by it. To identify the source of its strength it is useful to compare this model to a similar one identified in mice. The same theoretical ring-attractor has been suggested for head-direction representation in mice, but no corresponding physical ring has been found. While in the fly we find an ‘orientotopic’ organization of the neurons so that nearby directions are represented nearby in the ring (similarly, opposite directions are represented opposite on the ring), in the mouse no such organizing principle has been identified. This does not mean that scientists came away empty-handed. Instead, using a popular method of manifold analysis, they recorded the population activity of neurons in the anterodorsal thalamic nucleus that are modulated by head-direction, and conducted dimensionality reduction. Following this technique the authors identify a ring of stable activations, so that the position of neuronal activity on the ring corresponds to the head-direction of the mouse (Fig. 2). Here, the identified ring is completely abstract and not limited by physical properties; there is no principled way, prior to measurement, to predict the head-direction of the mouse from the activity of a single neuron, nor a principled way to predict how two head-direction neurons will relate to one another. Nonetheless, following this analysis, a variable extracted from the identified ring closely correlates with the head-direction of the mouse (Fig 2).

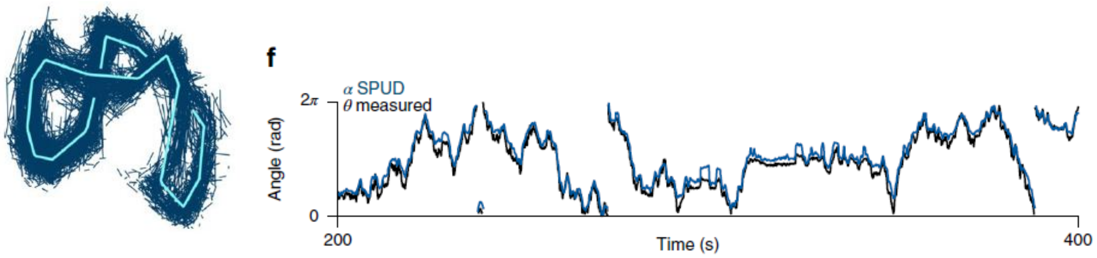

Figure 2. Adapted from (Chaudhuri et al. 2019, Fig. 2) with permission from Springer Nature; no further reproduction or distribution is permitted.). Left: a visualization of the ring of head-direction representation in a brain area called the anterodorsal thalamic nucleus of a mouse. Here, each point on the ring corresponds to some state of neuronal activity, but it is impossible to know just from the visualization what this state is. Right: a close correspondence between the prediction of head-direction from the abstract ring, denoted as α, and measured head-direction, denoted as θ.

By simply considering the two models it is hard to deny that the claim that scientists have identified head-direction representation seems much more convincing in the fly than in the mouse. Comparing the results and conclusions in both cases helps identify what makes the ring model in the fly stand out.

First, note that as explanations of the cognitive capacity of representation of head-direction, the cases of the fly and the mouse are relatively similar. Both describe a dynamical ring-attractor model. The main difference is the manner in which the function is implemented: localized and organized according to direction in the fly and distributed and abstract in the mouse. Both are viable neural implementations that can serve their causal role in this context.

However, when considering the explanation of neural properties, a major difference becomes evident. An explanation of the neural properties in the fly must explain not only why neurons are modulated by head-direction, but also why such neurons are localized in an anatomical ring, which is organized according to preferred direction. Alternative explanations are always possible, but it seems incredibly likely that the explanation is that this brain area has been historically shaped, also through selection, to take part in the representation of head-direction. It is difficult to imagine how such an organized anatomical structure would result otherwise. Therefore, it can reasonably be inferred that neurons in the ellipsoid body have the function of representing head-direction.

The neural properties in the mouse brain are not so easily explained. While one may certainly suggest that the sensitivity of some neurons to head-direction and the low-dimensionality of the population activity is because of a history of selection. This explanation still leaves many open questions: why are "head-direction neurons" dispersed among neurons with other sensitivities? Why these neurons and not others? How is the needed connectivity established in this population? And more. Additionally, because neurons are not localized or organized, one may wonder whether the analysis captured the relevant functional unit, opening the door to alternative explanations. To illustrate, it is possible that these neurons are involved in navigation as part of a larger population, and some correlate with head-direction as an indirect by-product. Some may worry that scientists are too quick to infer some specific history from limited data (Gould and Lewontin 1979). But this example demonstrates that not all cases lend themselves so easily to such inferences, and some cases are clearly more convincing than others.

Such an analysis is not possible if one adopts a systemic view of function. Recall that on this view to have a function is simply to have a causal role in some more general capacity that scientists aim to explain. But the properties that stand out about the fly case are not relevant in this context. Both in the mouse and in the fly, neurons can have a causal role in the representation of head-direction. Whether the neurons are organized or localized is not directly relevant to this causal role. Therefore, causal role functions cannot account for the intuitive difference between the fly case and the mouse case.

The upshot of this analysis is that, to use IBE from neural data, scientists should adopt a historical view of cognitive and neuroscientific function. Descriptions of mechanisms that ignore the etiological processes that shaped organization and activity in the brain cannot be supported with neural evidence through IBE.

One may worry that the case of the ring model in the fly is an extreme case that does not reflect how brains are generally structured. Indeed, it is very rare to find cases where functional properties map so nicely to anatomical ones. But this surprising correspondence between structure and function and its power in allowing scientists to draw clear inferences may be a useful inspiration for future scientific work. Instead of focusing on neural activity and the causal role it plays in cognitive capacities, scientists should pay more attention to neural properties that can serve as evidence about the history of the nervous system. Such properties include anatomy, morphology, gene expression, developmental processes, and more. Certainly, I am not the first to suggest that such properties are deserving of scientific attention, but the framework that this paper presents provides a clear reason: such properties can support models of cognitive capacities through IBE in a way that neural activity alone cannot.

5. Why IBE in neuroscience?

One can resist the conclusion of this paper by resisting the claim that data about neural activity supports models through IBE and suggesting instead alternative paths for this support. It is not the aim of this paper to argue for IBE as the right framework for interpreting the role of neural evidence. This has been done elsewhere (Calzavarini and Cevolani 2022; Coraci et al. 2024). Still, it is worth noting prominent alternatives and why I think they fail.

One could suggest something like a hypothetico-deductive account: models, with auxiliary hypotheses, make predictions about neural properties, and the degree to which such predictions come true is the degree to which the model is supported.

Putting aside the well-known philosophical difficulties with this account, it simply does not work in neuroscience. Making predictions about the zoo that is neural properties from cognitive models is incredibly difficult to currently do. As mentioned, very little is known about the form that the physical implementation of cognitive functions is expected to take. Scientists currently report findings about neural activity related to cognitive processes from the level of single spikes, through firing rates, activity in brain areas, and complex activity in networks. Thus, the physical implementation of cognitive models may have many different forms and it certainly cannot be deduced from a cognitive model.

A more flexible approach to confirmation is the Bayesian approach. Here theories need not deduce data, but simply predict it with some probability. An additional benefit is that, although Bayesian confirmation may be presented as abductive (Poldrack 2006), it has no requirement for a relationship of explanation between hypothesis and evidence. However, Bayesian approaches to confirmation suffer from the same difficulties as hypothetico-deductive approaches; there is no way today to determine the probability of some implementation according to a hypothesized cognitive model, and therefore no way to employ this approach. Current uses of Bayesian confirmation employ very rough proxies at best.

Moreover, often in neuroscience, null-results are not very informative: failing to identify neural implementation does not diminish support for a cognitive model, because scientists may have failed to examine the right kind of neural properties or the right location. The above approaches to confirmation require that null-results be informative about a model. In contrast, under IBE it is possible for neural results to allow for inferences when null-results do not.

6. Conclusions

I argued in this paper that if one accepts an IBE framework for the role neural evidence has in supporting cognitive models, then one cannot hold a systemic view of functions. For such functions cannot explain neural properties. Instead, one should adopt an etiological view of function, which naturally explains neural properties as the result of causal processes. To demonstrate how this can be done, the evidence for the ring-attractor model in the fly and in the mouse was described, where only the former leads to strong conclusions about function. The upshot is that there is benefit in adopting a historical view of functions and in aiming to build models that are informed by the system's history. There are also pragmatic implications to scientific research: scientists should dedicate more attention to neural properties that may not have current causal roles, but can serve as evidence about the history of the system, such as anatomy, gene expression, and developmental processes.

References

Aizawa, Kenneth. 2026. “The Evidence for Representation.” *Philosophy and the Mind Sciences* 7 (1). https://doi.org/10.33735/phimisci.2026.11962.

Anderson, Michael L. 2010. “Neural Reuse: A Fundamental Organizational Principle of the Brain.” *Behavioral and Brain Sciences* 33 (4): 245–66. https://doi.org/10.1017/S0140525X10000853.

Boone, Worth, and Gualtiero Piccinini. 2016. “The Cognitive Neuroscience Revolution.” *Synthese* 193: 1509–34.

Burnston, Daniel C. 2016. “A Contextualist Approach to Functional Localization in the Brain.” *Biology & Philosophy* 31 (4): 527–50. https://doi.org/10.1007/s10539-016-9526-2.

Calzavarini, Fabrizio, and Gustavo Cevolani. 2022. “Abductive Reasoning in Cognitive Neuroscience: Weak and Strong Reverse Inference.” *Synthese* 200 (2): 70. https://doi.org/10.1007/s11229-022-03585-2.

Chaudhuri, Rishidev, Berk Gerçek, Biraj Pandey, Adrien Peyrache, and Ila Fiete. 2019. “The Intrinsic Attractor Manifold and Population Dynamics of a Canonical Cognitive Circuit across Waking and Sleep.” *Nature Neuroscience* 22 (9): 1512–20. https://doi.org/10.1038/s41593-019-0460-x.

Coraci, Davide, Igor Douven, and Gustavo Cevolani. 2024. “Inference to the Best Neuroscientific Explanation.” *Studies in History and Philosophy of Science* 107 (October): 33–42. https://doi.org/10.1016/j.shpsa.2024.06.009.

Craver, Carl F. 2007. *Explaining the Brain*. Oxford University Press.

Craver, Carl F. 2013. “Functions and Mechanisms: A Perspectivalist View.” In *Functions: Selection and Mechanisms*, edited by Philippe Huneman. Springer Netherlands. https://doi.org/10.1007/978-94-007-5304-4_8.

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

Cummins, Robert. 2000. “‘How Does It Work?’ Vs. ‘What Are the Laws?’ Two Conceptions of Psychological Explanation.” In *Explanation and Cognition*, edited by F. Keil and Robert A. Wilson. MIT Press.

Fodor, Jerry A. 1968. *Psychological Explanation: An Introduction To The Philosophy Of Psychology*. Random House.

Fodor, Jerry A. 1974. “Special Sciences (Or: The Disunity of Science as a Working Hypothesis).” *Synthese* 28: 97–115.

Garson, Justin. 2017. “A Generalized Selected Effects Theory of Function.” *Philosophy of Science* 84 (3): 523–43. https://doi.org/10.1086/692146.

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. Series B. Biological Sciences* 205 (1161): 581–98. https://doi.org/10.1098/rspb.1979.0086.

Guest, Olivia, and Andrea E. Martin. 2023. "On Logical Inference over Brains, Behaviour, and Artificial Neural Networks." *Computational Brain & Behavior* 6 (2): 213–27. https://doi.org/10.1007/s42113-022-00166-x.

Jonas, Eric, and Konrad Paul Kording. 2017. "Could a Neuroscientist Understand a Microprocessor?" *PLOS Computational Biology* 13 (1): e1005268. https://doi.org/10.1371/journal.pcbi.1005268.

Kim, Sung Soo, Rouault Hervé, Shaul Druckmann, and Vivek Jayaraman. 2017. "Ring Attractor Dynamics in the Drosophila Central Brain." *Science* 356 (6340): 849–53. https://doi.org/10.1126/science.aal4835.

Maley, Corey J. 2018. "Toward Analog Neural Computation." *Minds and Machines* 28 (1): 77–91. https://doi.org/10.1007/s11023-017-9442-5.

Marom, Shimon, Ron Meir, Erez Braun, Asaf Gal, Einat Kermany, and Danny Eytan. 2009. "On the Precarious Path of Reverse Neuro-Engineering." *Frontiers in Computational Neuroscience* 3 (May). https://doi.org/10.3389/neuro.10.005.2009.

McCaffrey, Joseph B. 2023. "Evolving Concepts of Functional Localization." *Philosophy Compass* 18 (5): e12914. https://doi.org/10.1111/phc3.12914.

McCaffrey, Joseph B., and Jessey Wright. 2022. "Neuroscience and Cognitive Ontology: A Case for Pluralism." In *Neuroscience and Philosophy,* edited by Felipe De Brigard and Walter Sinnott-Armstrong. MIT Press.

Millikan, Ruth Garrett. 1989. "In Defense of Proper Functions." *Philosophy of Science* 56 (2): 288–302.

Neander, Karen. 1991. "Functions as Selected Effects: The Conceptual Analyst's Defense." *Philosophy of Science* 58 (2): 168–84.

Novick, Rose. 2023. *Structure and Function*. Cambridge University Press. https://doi.org/10.1017/9781009028745.

Piccinini, Gualtiero, and Carl F. Craver. 2011. "Integrating Psychology and Neuroscience: Functional Analyses as Mechanism Sketches." *Synthese* 183: 283–311.

Poldrack, R. 2006. "Can Cognitive Processes Be Inferred from Neuroimaging Data?" *Trends in Cognitive Sciences* 10 (2): 59–63. https://doi.org/10.1016/j.tics.2005.12.004.

Shapiro, Lawrence A. 2017. "Mechanism or Bust? Explanation in Psychology." *The British Journal for the Philosophy of Science* 68: 1037–59.

Taube, J. S., R. U. Muller, and J. B. Ranck. 1990. “Head-Direction Cells Recorded from the Postsubiculum in Freely Moving Rats. I. Description and Quantitative Analysis.” *Journal of Neuroscience* 10 (2): 420–35. https://doi.org/10.1523/JNEUROSCI.10-02-00420.1990.

Turner-Evans, Daniel, Kristopher T. Jensen, Saba Ali, et al. 2020. “The Neuroanatomical Ultrastructure and Function of a Biological Ring Attractor.” *Neuron* 108 (1): 145-163.e10. https://doi.org/https://doi.org/10.1016/j.neuron.2020.08.006.

Turner-Evans, Daniel, Stephanie Wegener, Hervé Rouault, et al. 2017. “Angular Velocity Integration in a Fly Heading Circuit.” *eLife* 6: e23496. https://doi.org/10.7554/eLife.23496.

Viola, Marco. 2021. “Beyond the Platonic Brain: Facing the Challenge of Individual Differences in Function-Structure Mapping.” *Synthese* 199 (1): 2129–55. https://doi.org/10.1007/s11229-020-02875-x.

Woodward, James. 2003. *Making Things Happen*. Oxford University Press.

Zhang, K. 1996. “Representation of Spatial Orientation by the Intrinsic Dynamics of the Head-Direction Cell Ensemble: A Theory.” *Journal of Neuroscience* 16 (6): 2112–26. https://doi.org/10.1523/JNEUROSCI.16-06-02112.1996.