# Interview with Philosophie Magazine, 6th January 2026

**https://www.philomag.com/articles/mazviita-chirimuuta-les-modeles-neuroscientifiques-doivent-etre-interpretes-avec**

**N1**

**Are the models of the brain that are offered to us from the neurosciences too simplified to be honest ? That's the lesson of humility suggested by *The Brain Abstracted* (*MIT Press, 2024*) which has received the Lakatos prize following the prize from the *Royal Institute of Philosophy.* The author, Mazviita Chirimuuta, who teaches at Edinburgh University, has agreed to answer our questions.**

N2

**In what sense is the image of the brain stemming from neuroscience deceptive ?**

There is a famous saying by the statistician George Box : « all models are wrong, but some are useful ». People who develop and use scientific models are well aware that they must depart from many of the known facts of the things they represent, in order for their models to serve as useful representations. However, in the image of science presented to the general public, models and other scientific representations are taken to be clear windows on reality – just showing us how things are. I think this is particularly worrying when models which assume that the brain is a computer, like an artificial neural network in AI, are presented to the public as a literal reality even though they depart from many of the known facts. My central argument in the book is that neuroscientific models need to be *interpreted* more carefully and not taken literally. I'm not trying to argue that neuroscientists themselves should build different kinds of models, less deceptive ones. I like to paraphrase Karl Marx: "my task as a philosopher is to interpret neuroscience, not to change it!"

**But how can we know if the image that neuroscientists present of the brain is reductive or simplifying ? What point of comparison allows you to affirm this ?**

This is a question I am often asked : "if my thesis is that science *cannot* represent the brain in its full complexity, then how can we even know that there is this complexity that is being left out from neuroscientific theories and models?" In the book I offer a variety of ways around this apparent obstacle. Through historical case studies of theoretical neuroscience before and after the introduction of the computationalist framework, we can see the decision points at which scientists have opted for more radical simplifications. For example, in the late nineteenth and early twentieth century, physiologists such as Charles Sherrington and psychologists such as Ivan Pavlov attempted to reduce all of the activity of the nervous system to the operation of simple reflexes. We learned from subsequent research that this was a vast oversimplification. However, I argue, there is a continuity in the ethos from the era of the reflex theory to the era of computational neuroscience, via the cybernetics movement of the mid twentieth century. In computational neuroscience it is assumed that neurons operate as simple input-output calculating devices, like electronic components. From neurobiology – the experimental research that uses methods prevalent in other kinds of cell biology - we learn that neurons are vastly more complicated than that, and that many of their anatomical and chemical properties are crucial to their operations, as well as their interactions with other non-neuronal cells. Yet, neurocomputational models ignore most of these details. As I said in my previous answer, this does not undermine their utility as scientific models, but it does force us to question computationalism as a metaphysics of the brain and mind. And this metaphysical position has many

current ramifications: computationalism is the leading philosophy in silicon valley and amongst transhumanists who believe that the mind can be 'uploaded' into a computer.

**You explain that we cannot understand the brain via simple models. But isn't it always the case that we begin to understand by simplifying or at least modelling reality ?**

Yes, in fact I think there is a deep connection between simplification and scientific understanding. So really I don't say that we *fail* to understand the brain through simplification because I think scientific understanding *consists* in the production of these simple theories and models, as well as production of models systems, experimental objects that remove the interfering factors that occur in the natural environment. One of the slogans of my book is: *scientists make things simple by making simple things*. The more radical conclusion is that the brain 'as it is in itself', i.e., in its full complexity, is not comprehensible to science.

**Mightn't it be necessary to consider neuroscience as a science still in a state of infancy…. what benefit would there be of this show of humility ?**

I don't think it's helpful to attribute the current state of neuroscience to its immaturity. Many of the methods employed within neuroscience have been perfected in other branches of science and have a long history of use there. For example, in Chapter 7 I discuss the dynamical systems approach which comes from physics and is now getting popular as a method for modelling the brain. The more important issue, in my opinion, is not that physics is more 'mature' but that it deals with inherently simpler objects. Samples of non-living matter can be made perfectly homogeneous, whereas no two cells are ever identical, let alone any two mice (even if genetically identical). Non-living matter is fairly insensitive to its surrounding environment. If kept under the same laboratory conditions you can expect a sample of iron to behave in exactly the same way, anywhere in the world. In contrast, the brain and nervous system are exquisitely sensitive to surroundings – as they must be, for without this we could not perceive and respond to the world around us. Also, neuronal responses are modulated by the whole life history of the organism, which is how memories are inscribed in them. In this chapter I argue that the fundamental changeability of the brain cannot be represented within the dynamical systems framework, even though its purpose is to represent the change of state of a physical system. Again, this does not undermine its utility; it is just to say that there are important properties of the brain that such models necessarily exclude.

**Must we give up the attempt to understand the brain or can we still understand it but in a different way ? In that case, how ? How can the functioning of the brain be represented ?**

The agenda of my book is not to offer neuroscientists a different plan of action, but to say how philosophers and other interested parties should interpret actually existing neuroscience. As I said, I think that current neurocomputational and dynamical systems models do provide understanding of the brain, but they do not tell the whole story. It is interesting to consider that there may be alternative perspectives that can offer insights into what those frameworks leave out (while probably introducing their own blind spots). Since publishing the book I've been looking more into biological theories of cognition, ones that treat 'information processing' as a generic feature of living cells. A book by Peter Sterling and Simon Laughlin, *Principles of Neural Design,* first made me aware of this idea. A new book by the neuroscientist Nicole Rust, *Elusive Cures*, also considers the problem of over-

simplification in neuroscience, especially in research aiming at cures for brain diseases, and proposes some new ideas based on complexity science. I would highly recommend that book as well.

**What is the « haptic realism » that you defend ?**

*Haptic realism* is an account of scientific knowledge, loosely inspired by Kant. It asserts that scientific knowledge is the product of an interaction between researchers and the items they investigate. The active contribution of the scientist cannot be discounted, meaning that the standard of traditional scientific realism, that science at its best should represent things as they are in themselves, independently of the way that humans have chosen to interact with them, is an unobtainable ideal. Haptics is active touch and I propose this as metaphor for the way that scientific research, and hence the resulting knowledge, involves deliberate manipulation and shaping of the objects investigated, in order that the practical goals of research, such as medicine and technology, can be reached.

Once it is recognised that science is not a disinterested pursuit of knowledge for knowledge's sake, but an activity directed at producing knowledge that is utilisable, it is easier to see how it is possible that some of the most prized products of scientific research can be full of idealisations and other simplifications that depart deliberately from the truth of observable empirical facts. The point here is that idealisation is a method for taming complexity, and complexity is more often than not an obstacle to practical efficacy. Science conceptually and materially shapes its objects with a view to isolating the causal relations most of relevant to human ambitions of manipulation and control. I have spoken here of science in general. I think these points apply not only to neuroscience but the also the other biological sciences, these all being connected in some way to biomedical goals even when classified as basic research, not applied science. Most of the physical sciences can be understood as having connections with application, however indirect, though there may be exceptions such as cosmology.

When writing the book I was particularly interested in neo-Kantian philosophy of science, especially the work of Ernst Cassirer. One could also find close parallels, perhaps closer, with pragmatist philosophy of science, especially the account of John Dewey (1929) in *Quest for Certainty.* A new pragmatist version of realism, quite similar to haptic realism, is presented by Hasok Chang (2022) in *Realism for Realistic People*.

**Are you inspired by Henri Bergson who already a century ago denounced the tendency to reduce consciousness to the brain ?**

I became interested in Bergson because of my reading of Cassirer. Cassirer's major theoretical work, *Philosophy of Symbolic Forms, volume 3 (Phenomenology of Cognition)* is in many ways a response to Bergson. I found myself very sympathetic to Bergson's views, presented in *Evolution Creatrice*, on the limitations of scientific abstractions, especially when representing living things: the need to reduce change to stasis, and the tendency to ignore the particularity and heterogeneity of living organisms. Before publishing the book I wrote a paper comparing the ideas of Bergson, Husserl, Canguilhem and Merleau-Ponty on scientific abstraction, showing how in the work of these last two philosophers there was an incipient criticism of computationalism. This was in their writings from the 1960s on cybernetics. So Bergson was certainly an inspiration but I have to admit that I became more familiar with this philosophy of science than his philosophy of mind. *Matière et mémoire* presented me with an interpretative challenge that I had to postpone to a later date!

**What place do you give to the « mind » ?**

In the book I don't offer a positive account of the mind and its relationship to the brain. I do make a negative case against neuro-reductionism, the idea that the best way to learn about the mind is to study the brain. My point is that because of the simplifying strategies that cognitive neuroscience must employ when, for example, performing experiments involving tasks such as memory and decision making, we should not assume that the discoveries in the lab tell us much about the behaviours we are concerned with in actual human lives. I see this argument as a corrective to a trend that was previously quite dominant in the philosophy of neuroscience, called *neurophilosophy.* This was the idea that we could be reductionists and even eliminativists about mental concepts, replacing them with neuroscientific terms. There was no place for an autonomous psychology of the mind, let alone a philosophical enquiry independent of neuroscience. I have always disliked this kind of scientism – this imperialism of one form of enquiry. The mind is a universal topic of concern, and it can be investigated in countless ways, not restricted to scientific methods.

**Your book has been widely praised by the philosophical community… but how have neuroscientists reacted?**

I've been very pleased, and perhaps surprised, by the warm reception my book has had amongst neuroscientists who have read it (at least the ones who have spoken with me about it). For example, I have been interviewed on neuroscience podcasts ('Brain Inspired' and 'Theoretical Neuroscience', both highly recommended) and invited to some conference sessions. I think that the issue my book addresses is one that many neuroscientists worry about, but the pressure of laboratory research does not allow for systematic reflection. One role for philosophy of science is precisely to provide a forum for methodological discussions that the scientific treadmill tends not to allow.

**You began your studies in philosophy and psychology before turning towards neuroscience. Why the switch ?**

I must confess, in my youth I was a reductionist! I actually began my studies at university in physics and philosophy. I switched to a psychology and philosophy degree because I was more interested in minds and people than electrons and magnets. But I missed the precision and satisfying explanations that physics had provided. Then I discovered neuroscience, specifically, the branch of neuroscience that involved building mathematical models of cell responses in the visual system. That seemed to offer the best of both worlds: theoretical rigour and the promise of an explanation of the mind and brain.