Human Arenas
https://doi.org/10.1007/s42087-024-00423-4

RESEARCH

Reflections on Simplicity and Complexity in Computational Neuroscience

Kanaka Rajan[1,2] · Alessandro Treves[3] · Rocco Gaudenzi[4]

Received: 30 March 2024 / Revised: 12 June 2024 / Accepted: 13 June 2024
© The Author(s) 2024

Abstract
This double interview with two distinguished researchers in computational neuroscience, Kanaka Rajan and Alessandro Treves, aims to capture a part of their talks and discussions that emerged during a workshop on physical modelling of thought, held in Berlin in January 2023. The topic is the fascinating all-round intersection of physics and neuroscience through the perspectives of the interviewees. The dialogue traverses the complex terrain of modelling thought processes, shedding light on the trade-off between simplicity and complexity that defines the field of computational neuroscience. From the early days of physics-inspired brain models to the cutting-edge advancements in large language models, the interviewees share their journey, challenges, and insights into the modelling of physical and biological systems; they recount their experience with computational neuroscience, explore the impact of large language models on our understanding of human language and cognition, and speculate on the future directions of physics-inspired computational neuroscience, emphasising the importance of interdisciplinary collaboration and a deeper integration of complexity and detail in modelling the brain and its functions.

Keywords Modelling practices in physics and biology · Computational neuroscience · Physics modelling of thought · Large language models

Introduction

One way to answer the question of what is the gist of modelling is to say that modelling consists in striking a balance between simplicity and complexity. A good model is complex enough to include all the necessary features and simple enough to leave out the superfluous

✉ Rocco Gaudenzi
rocco.gaudenzi@univr.it

1 Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA

2 The Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA

3 SISSA - Cognitive Neuroscience, Trieste, Italy

4 Department of Human Sciences, University of Verona, Verona, Italy

Published online: 21 June 2024

ones, not in absolute terms, but in relation to the specific purpose of the model. From Galilei onwards, physics has tackled this issue, and inevitably shaped the idea of modelling in other fields of the natural and social sciences. A representative of this influence is computational neuroscience, where a range of modelling strategies are adopted with the purpose, broadly speaking, of modelling thought. This article-interview aims at collecting the views of the two physicists turned computational neuroscientists Kanaka Rajan and Alessandro Treves on the intersection of physics and neuroscience. Originating from a workshop held at the Max Planck Institute for the History of Science, Berlin, in January 2023, the conversation explores the evolution, challenges, and insights of modelling thought processes and physical systems in computational neuroscience. Both researchers share their transitions from physics to biology, emphasising the role of complexity and biological details in understanding neural systems. They discuss the limitations and potentials of current computational models, including large language models, in capturing the essence of human cognition, language, and emotion. Rajan highlights the importance of integrating biological complexities into models, moving beyond the simplicity prized in physics to embrace the 'messiness' of biological systems for a more comprehensive understanding. Treves offers a critical view on the fragmented state of computational neuroscience, and reflects on the sociological impact of physics-inspired approaches to neuroscience, questioning the concrete advancements, but acknowledging the interdisciplinary bridges built between fields. The heterogeneity of answers in terms of style and focal points, partly due to the fact that the interviewees are at different points in their career, contributes to drawing a multifaceted picture of the interdisciplinary field of computational neuroscience.

Interview

Rocco Gaudenzi

On a personal level, what is the main difference you encountered between the modelling of physical systems and biological systems? Was it a challenge to shift from a physics-modelling mindset to modelling biological systems?

Kanaka Rajan

Absolutely, it was a challenge. Computational neuroscience initially leaned towards simplifying problems, approaching the brain with the same attitude as any other physics and engineering problem. This early focus had to gradually shift to accommodate the critical role of biological complexities. The transition from physics-based modelling to biological systems is not linear. In physics, we often strive to simplify problems and search for the most elegant, often mathematical, solution. We treat noise as something to be eliminated. However, in biological systems, particularly in computational neuroscience, which was largely pioneered by physicists and engineers, the approach differs fundamentally.

The main difference lies in how we deal with complexity and details. In biological systems, you cannot just 'clear the deck' of details as these details are crucial. They are not just noise; they are integral to understanding the system. In physics, we are taught to seek simplicity and generalizability. But when you shift to biology, you realise that what we once considered 'noise' or extraneous details are actually carrying out most of the work. Unlike in physics, where a singular, elegant solution is often sought, we need many models in biology. Each model might be 'right' in its own way, reflecting the intricate nature of biological systems and brains. The shift requires a fundamental change in mindset: from seeking to eliminate details, to understanding and incorporating them as vital input into a complex set of models. This is the core of the tension in moving from physics to biology modelling.

Alessandro Treves

By sheer chance, I was in Jerusalem in the fall of 1987 to witness, almost as a neutral observer, the formation of an international community interested in the brain from a statistical physics perspective. For my PhD I had moved to Israel, out of curiosity, with a vague idea to keep studying fundamental physics, as long as it continued to be aesthetically pleasing. On one side, string theory proved beautiful, but to me rather inaccessible; on the other side, there was excitement on campus about the "solution" of the Hopfield model of associative memory by Amit, Gutfreund and Sompolinsky (1985). So I was tempted to transition. The turning point was their organisation, in '87-'88, of the year-long workshop on Physics and the Brain, essentially a crash neuroscience course for passionate but ignorant physicists of different ages. Researchers came to participate, for periods of variable durations, from all over Europe and the rest of the world, and it was an extraordinary and empathic display of naiveté and enthusiasm. Simple questions about brain function were posed or recast in the basic language of Ising models of disordered systems, the only *koiné*[1] for most of the attendees. It is true that, particularly at the final meeting, a 'Bat Sheva' conference with many participants from Europe and North America, some also came who were engaged in more realistic, biologically sound modelling of neural systems. They were non-physicists mostly from North America—for reasons internal to the condensed matter community, physicists there had been left out of the buzz around spin glasses, centred in Europe—and their modelling work appeared to my student eyes much more of a run-of-the-mill, almost technical exercise, devoid of major intellectual ambition. The 'real thing' had to involve a leap of faith: Physics. *Credo quia absurdum*—I believe in what I do not understand.

Rocco Gaudenzi

Is it more fruitful to have a biologically detailed model of the brain—i.e., informed by brain activity and connectivity—which possibly explains and reproduces aspects of our higher-level behaviour, or rather a simple abstract model—e.g., simple neurons and connections with no non-linearities? What are your personal inclinations?

Kanaka Rajan

My personal inclination, shaped by my training in the physical sciences, initially leaned towards simpler, more abstract models. The elegance of simplicity, a tool that suffices with minimal complexity, was my go-to approach. However, through my interactions with experimentalists, I have learned that the simplest models often fall short, leading me to appreciate

[1] The word indicates the Greek language spoken by various peoples from the close of Classical period to the Byzantine Era. Today it is used as a synonym for *lingua franca*.

the intricate dance of biological details and their impact on our understanding of brain activity. Collaborating with the experimentalists who are in touch with the 'ground truth' of real brain data, significantly evolved my perspective.

I quickly realised that these simple, abstract models hit a wall when confronted with the complexities of biological systems. In physics, we might disregard certain details as noise, but in neuroscience, these details are fundamental to our understanding. The intricate interplay of these details significantly influences our interpretation of brain activity and connectivity.

For example, our work on multi-region 'network of networks' models has shown us that simple neural network models often fail to capture the complex interactions across different brain regions. By incorporating detailed network dynamics, our models better explain brain-wide interactions and their role in cognitive function. Similarly, our work on multitask learning illustrates that simple, monolithic models are not sufficient for understanding how the brain handles multiple tasks, and suggests that modular, hierarchical structures are necessary to capture the flexibility and efficiency of biological learning processes.

Now, my approach is more nuanced. I am seeking a balance, integrating the simplicity of the models I am familiar with and layering them with biology's complex, 'gnarly' details. It is not about taking everything at face value, but discerning which details are crucial and which are not. This approach, a blend of physics-inspired simplicity and the rich complexity of biology, is what I believe will lead to a deeper, more comprehensive understanding of the brain. It is a journey from simplicity to complexity, guided by the insights gleaned from experimental data and the inherent messiness of biological systems.

Alessandro Treves

In its early days, abstract physics-inspired modelling of the brain served mainly as mass psychotherapy for a group of highly intelligent individuals who found, there, new meaning to repurpose their endeavours. They had been attracted by the odd properties of spin glasses—usually without having ever seen a spin glass, nor knowing whether they actually existed (Mezard 2022)—and had been initiated to the mysteries of the replica method, but then were beginning to wonder whether it would lead anywhere, beyond endless variants of abstract model "solving", which had started to feel like scientific onanism. Accessing the brain released that energy in a variety of potentially fruitful directions. With hindsight, has it been that fruitful? Nearly forty years later, it is often said that concrete progress directly arising from the analysis of models based on the physics of disordered systems has been limited. I do not quite agree, but it is fair to admit that the main effect has been sociological, in facilitating access to neuroscience initially for some, later for many more, who had been attracted by physics in their youth; and, conversely, in giving access to their way of thinking to many neuroscientists with a biology or psychology background. For me personally, bridging and combining conceptual perspectives has been a core component of the pleasure. So, the issue has never been defending a disciplinary boundary, or claiming victory for one side; rather, to trespass and eventually erase the boundary, not only between physics and biology, but also with psychology, linguistics and other communities, by trying to internalise their diverse narratives.

Rocco Gaudenzi

Is the final goal of computational neuroscience a realistic (i.e., biologically detailed) modelling of the brain?

Kanaka Rajan

The essence of computational neuroscience is not solely about realistic modelling. It is deeply rooted in the quest to understand the brain and its complexities. While realistic modelling is a valuable tool in this journey, it serves as a means to a larger goal: a more profound understanding of the brain and its functions.

This kind of detailed, realistic modelling is increasingly within reach. We are on the verge of a significant breakthrough with the creation of the first complete connectomics map of a mammalian brain, a step that promises to revolutionise our approach to understanding brain activity. Such detailed maps are vital, but they represent just one piece of the puzzle. The real challenge lies in discerning which details from these maps matter for understanding cognitive processes like memory and decision-making.

In computational neuroscience, it is not just about accumulating every minute detail; it is about identifying which aspects of biological complexity are necessary for understanding and which can be abstractly represented. As the field advances, so do our models and theories, continually redefining the limits of our understanding. We are part of a continuous journey of discovery, with each advancement pushing the boundaries of what we thought possible. This field has boundless potential. With each new complexity we investigate, we uncover more messy details we do not yet understand. Our work is constantly evolving to meet the expansive and retrospective understanding of complex systems.

Alessandro Treves

I do not see a widely shared research program of computational neuroscience. There are organised ‘big science’ campaigns to obtain funding, which span the continuum between exercises of public relations and outright swindles, and there are trendy computational approaches, usually domain-general, that in the very opinion of some of their leading figures have generated very little theoretical understanding. There are, on the other hand, a variety of bold individual attempts to understand aspects of neural computation, often carried out in conjunction with brilliant experiments, and which tend to make advances inversely proportional to the degree to which they were predictable and classifiable within a well-defined research program as well as to the number of coauthors.

Rocco Gaudenzi

Today it seems possible to imitate our linguistic behaviour by means of large language models which feature a remarkably conceptual and constructive simplicity and abstractness. What are the implications of these results for understanding our own behaviour and abilities?

Kanaka Rajan

The advancements in large language models, particularly their ability to imitate linguistic behaviour, offer a fascinating window into the complexity of human language and cognition. These models, with their conceptual simplicity and ability to construct coherent responses, demonstrate the power of probabilistic modelling. They are capable of producing language that resonates with human communication patterns, but there is a significant gap between mere production and deep understanding. This advancement in language models highlights a crucial distinction: while we can mimic the structure of language, understanding the rich tapestry of human cognition, culture, and emotion remains a vast frontier to explore.

The key implication here is the distinction between generating language and truly grasping the nuances of human thought and culture. Current models can replicate patterns and even adapt their outputs based on context, which is impressive. However, this is not the same as understanding. Understanding involves grasping semantics, culture, and the subtleties of human interaction, that these models are still far from achieving.

We are at a point where the ability to mimic language opens up questions about the nature of understanding itself. It is one thing to produce a string of coherent words; it is another thing to capture the essence of human thought, motivation, and emotion that underlies those words. This gap highlights the complexity of our brains and the intricacies of our cognitive processes.

So, while these models are a leap forward in computational linguistics, they also underscore the vastness of what we have yet to understand about our own brains. They remind us that human cognition is not just about processing information, but about integrating it into a rich tapestry of cultural, emotional, and experiential contexts.

Alessandro Treves

Large language models are having a profound impact on many aspects of our lives, much more dramatic than their contribution to understanding human language. Conceptual understanding, however defined, is not their goal. In a sense, their main contribution is to clarify which aspects of natural language can be handled with massive computing power, extraordinary amounts of data, and no insight about brain mechanisms. This is a valid contribution. They have little to say about other aspects, such as the time course of language acquisition in children and adults, or what has produced the diversity of syntactic structures observed around the world, or about what characterises the forces driving language evolution, or how language can be harnessed in creative and innovative thinking. There might be progress in the near future derived from a combination of machine learning and classical linguistic techniques, but so far it has not been too visible.

Rocco Gaudenzi

Could we reproduce and explain our linguistic behaviour—how we talk, why we give a certain answer, say a word and not another one, etc.—with a model?

Kanaka Rajan

The question of reproducing and explaining our linguistic behaviour with a realistic model is layered and complex. In some respects, yes, we can reproduce aspects of linguistic behaviour. Current models can generate language that mimics human speech patterns to a remarkable degree. They can, to an extent, choose words and construct sentences based on context and probability distributions. But this is reproduction, not explanation.

Explanation requires a deeper understanding of not just the mechanics of language, but also the underlying motivations, emotions, and cultural contexts that shape why we say what we say. It is about bridging the gap between the structural aspects of language and the rich, often messy, tapestry of human cognition and experience.

Currently, our models can mimic the pattern of speech, but they lack the depth of understanding that comes with human experience. They do not grasp the why behind our words. For instance, the difference between someone authentically feeling an emotion and an actor convincingly portraying it is a matter of internal experience versus external expression. Our models are akin to skilled actors—they can reproduce the external expressions of language, but they lack the internal cognitive and emotional context.

To truly explain our linguistic behaviour, we need to go beyond the surface. We must delve into the complexities of human cognition, emotion, and culture. This is a vast, uncharted territory where computational models have only begun to scratch the surface. So, while we have made strides in reproducing linguistic behaviour, the journey to fully explaining it is much longer and more intricate.

Alessandro Treves

Anybody who has reviewed admission essays by applicants for a PhD program will concur that ChatGPT explains, reproduces and also improves the linguistic behaviour of the root-mean-square candidate, and thus provides a useful filtering-out mechanism if one seeks candidates who are ChatGPT-irreproducible. So, we do not need a realistic model for that. When it comes to non-ordinary linguistic behaviour, for example the production of great poets, reproducing its uniqueness (assuming it not to be a self-contradictory concept) seems a long way ahead, even with large language models. It is not clear whether using a realistic model—the question is what is the sense of realistic?—would yield any advantage, and it would probably further obscure any significant understanding of how creativity may work.

There is however a wide gap between large language models from artificial intelligence and realistic models that reproduce neuronal biophysics to some detail. More specifically, there is a large space for abstract models that, unlike large language models, attempt to address neural mechanisms while taking into account key constraints of the human brain. Significant progress is likely to arise from within this space, which is relatively independent of both big science and big data.

Rocco Gaudenzi

Where do you think that physics-inspired computational neuroscience will head to in the future?

Kanaka Rajan

The trajectory of physics-inspired computational neuroscience is positioned to take some exciting and transformative directions in the future. One of the most intriguing aspects will be the further integration of physics concepts and tools that have not yet been fully explored in the context of neuroscience, like the idea of embodiment and the impact of physical laws on our cognitive processes.

The notion of embodiment as a marriage between physics and neuroscience is fascinating. It is about acknowledging that our brains are not just abstract information processors; they reside in physical bodies that interact with the world according to the laws of physics. This interaction influences our cognitive processes in ways that have been largely overlooked in traditional computational models. By incorporating the principles of physics that govern our bodily interactions with the environment, we can develop a more holistic understanding of the brain.

The future of computational neuroscience might see the infusion of analysis tools derived from physics, providing novel insights into the functioning of neural networks and the brain's architecture. It is about looking beyond the neuron-to-neuron connections and understanding how the brain's physical structure influences its function.

In essence, physics-inspired computational neuroscience will likely continue to evolve, integrating more complex and nuanced understandings of how our physical existence and the laws governing it structure and constrain our cognitive processes. Recognizing these physical laws as fundamental constraints on cognitive processes offers an exciting frontier that promises to deepen our understanding of the brain in ways we have yet to fully imagine.

Alessandro Treves

I would prefer to call it neural computation rather than computational neuroscience. I expect creative researchers to be able to guess where the bulk of their colleagues are heading, so as to move in other directions. It is indeed surprising, in fact, how little some of the fundamental neural computation questions have been considered by the prevailing trends in the field. One classical example is neocortical lamination, which has been there for all to see since at least the days of Ramon y Cajal. It is a major feature in the organisation of mammalian cortices and yet has received little or no attention by neural computation researchers. How come reptilians have their cortical cells clustered in one layer, while we have three? All of us mammals, including those of us with a less intimate relation to their mothers, who have laid them as eggs. Another is odour representation and coding in the different stations of the olfactory system, the oldest to utilise cortical processing, and yet heroically resistant to the bulldozers that would flatten it, too, into a low-dimensional variety. Do the 'native' inhabitants of the cortex, the olfactory neurons, with their 1000-plus-dimensional activity space, express a richer variety than the visual, auditory, somatosensory neurons that later came to colonise their territory? A third riddle is the logic in the organisation and refinement of the basal ganglia in mammals, structures humiliated and made redundant by those narrowing it down to binary decision making. What is the exquisite balancing act they perform, whose importance we tend to only appreciate when they are off balance?

And dear to me, the characterization of phases and phase transitions with abstract mathematical models, which I am still pursuing—luckily in collaboration with outstanding students (Ryom & Treves, 2023). It has certainly gradually reshaped our intuition about memory attractors, but it appears not to have extinguished its potential contribution as a tool that, by converting the quantitative into the qualitative, completes the cycle begun with the proper quantitation of qualitative approaches to natural intelligence.

Author Contributions K.R. and A.T. answered the questions posed by the interviewer R.G., who also wrote the introduction.

Funding The research mentioned by A.T. was in part funded by HFSP Research Grant P0057/2016. R.G. and K.R. declare no funding.
Open access funding provided by Università degli Studi di Verona within the CRUI-CARE Agreement.

Data Availability No datasets were generated or analysed during the current study.

Declarations

Ethical Approval Not applicable.

Competing Interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Amit, D. J., Gutfreund, H., & Haim Sompolinsky. (1985). Storing infinite numbers of patterns in a spin-glass model of neural networks. *Physical Review Letters*, *55*, 1530.

Mézard, M. (2022). Spin glasses and optimization in complex systems. *Europhysics News*, *53*, 15–17.

Ryom, K. I., & Treves, A. (2023). Speed Inversion in a Potts Glass Model of Cortical Dynamics. *PRX Life*, *1*, 013005.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.