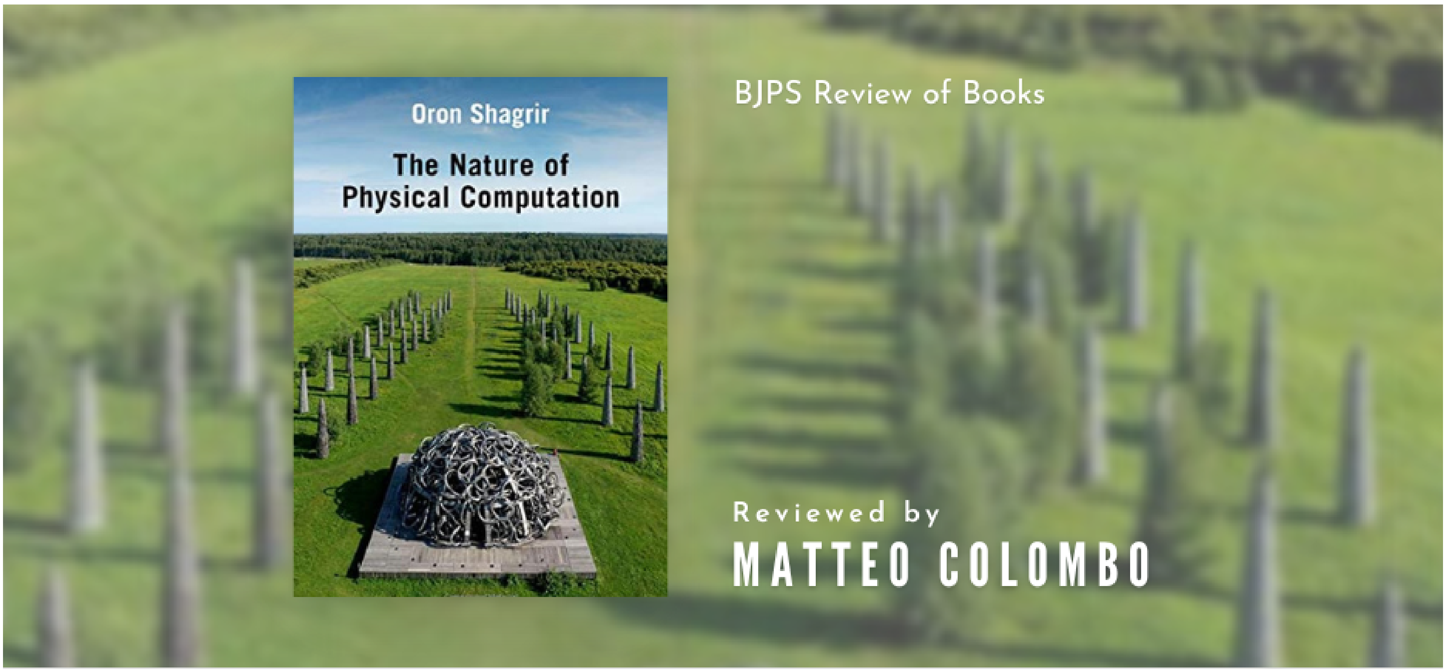

THE NATURE OF PHYSICAL COMPUTATION

Oron Shagrir

Reviewed by Matteo Colombo

The Nature of Physical Computation
Oron Shagrir
Oxford University Press, 2022, £47.99
ISBN 9780197552384

Recently, a computer engineer claimed that the LaMDA (Language Model for Dialogue Applications) chatbot that he helped to build had become conscious, seemingly able to think like a human. The engineer's company disagreed and put him on leave, claiming that there was plenty of evidence that that computing system was neither conscious nor thinking.

This event might strike you as perplexing, but it motivates questions lying at the core of the computational sciences. For example, what does it mean to say that a system like LaMDA computes but a system like the Apennine Mountains does not? Do nervous systems compute? If they do, in virtue of what do they compute? Do they compute in the same way as LaMDA does? And what is the relationship between computing and having a conscious mind?

Oron Shagrir's *The Nature of Physical Computation* illuminates these and other questions about the conceptual foundations of the computational sciences. It develops a novel account of physical computation that foregrounds the roles of computational modelling in the sciences of mind and brain, and whose central claim is that semantic properties are essential to individuating physical computing systems. Accordingly, LaMDA, my laptop computer, and perhaps my nervous system are systems that compute, partly because of their semantic properties; the Apennine Mountains, my chair, and perhaps my circulatory system are not computing systems because they do not have semantic properties.

The nine chapters in Shagrir's *The Nature of Physical Computation* clarify these ideas, by weaving together insights from computer science and logic, philosophical arguments about computation, mechanism, scientific modelling and representation, and detailed case studies from artificial intelligence and computational cognitive neuroscience. They do so in a no-frills, pellucid way, which makes this book a welcome addition to the burgeoning philosophical literature on physical computation. Compared to another recent book covering similar themes, Gualtiero Piccinini's ([2020]) *Neurocognitive Mechanisms*, Shagrir's book puts more emphasis on computational modelling and its methodological roles in the sciences of mind and brain, paying relatively less attention to the metaphysics of mind and the history of computing.

The argument that Shagrir develops for his modelling account of computation has a negative and a positive part. The negative part consists in rejecting two 'dogmas'. The first is the logical dogma that mathematical theories of computation in logic and computer science should ground any adequate account of physical computing system. The second is the architectural dogma that physical computing systems differ from non-computing systems in virtue of their abstract, causal (or functional) structure. The positive part leverages three ingredients. The first is that a physical computing system must implement a dynamical formalism of some kind, not necessarily a formalism from computability theory or logic. The second ingredient is that a physical computing system must possess some features that represent objects or properties in a target domain. The third and final ingredient is that the processes and representing features of a physical computing system must mirror certain processes and features in a target domain.

The book begins by laying out and motivating three general desiderata for an account of physical computation. According to the classification desideratum, an adequate account of physical computation should distinguish computing from non-computing systems, and should also help us to taxonomize distinct kinds of computing systems. The objectivity desideratum is the conjunction of two ideas characterizing a notion of partial objectivity (PO): (PO1) every computational property of some physical computing system is observer-independent (or not a matter of mere subjective interpretation), and (PO2) some computational properties of every physical computing system are observer-independent (or not a matter of mere subjective interpretation). The third desideratum, the utility desideratum, is that an account of physical computation should do justice to explanatory and methodological practices in the computational sciences by illuminating the role, relevance, and fruitfulness of applying computational descriptions to some systems but not others.

Notice that while the classification and utility desiderata are widely accepted and plausible, the objectivity desideratum is particularly contentious. To illustrate PO1, Shagrir refers to minds and brains, whose computational properties would not be a matter of interpretation. But, as acknowledged by Shagrir, it is not obvious that minds and brains are computing systems, and nor is it obvious that if they are computing systems, then this must be a presupposition rather than a matter of discovery. If the only illustration of PO1 is minds and brains, and whether minds and brains actually compute is contentious and a matter of ongoing inquiry, then PO1 may well be false. Intuitively, it does look as though familiar computing systems like laptops illustrate PO2, however. But here Shagrir argues that the architectural profile of a given artifact is not essential to its computational nature, while also acknowledging that the data structures in 'conventional computing systems' are representational only in virtue of how they are interpreted by their designers and users. So, it is also unclear whether PO2 is generally true of 'conventional computing systems', whose design is grounded in ideas from computer science that are reflected in their architectural profiles. Shagrir says, 'we should be open to the possibility that computational properties are not objective at all' (p.

25), but does maintain at least a weak objectivity desideratum in terms of PO1 and PO2 (for accounts denying objectivity desiderata, see Schweizer [2019]; Colombo [2021]).

Chapters 2 and 3 introduce and argue against the logical dogma. After explaining Alan Turing's and Alonzo Church's notions of effective computability, and situating them in their historical context, Shagrir argues that these notions are too restrictive to apply to physical computing. The upshot is that we should not apply computer scientists' notions of Turing machine, effective procedure, or algorithm in grounding an account of physical computation. Any kind of dynamical formalism—that is, any kind of mathematical structure describing the temporal evolution of a system—can be implemented by physical systems that compute.

Chapters 4, 5, and 6 examine the architectural dogma. Using Robert Cummins's ([1989]) account of computation as step-satisfaction as a foil, Shagrir argues that architectural features play no essential role in individuating physical computing systems. Shagrir accepts David Chalmers's ([2011]) notion of implementation of an abstract dynamical formalism by a physical system as a necessary condition on an adequate account of physical computation. Here, implementation involves a structure-preserving mapping between the formalism and the physical system, a grouping of states of the physical system into state-types, and the transitions between these states are causal relationships (pp. 130–31). But these features do not suffice for physical computation. Mechanistic features like a system's teleological function to compute and the vehicle of computation being medium-independent do not suffice either. The two extra ingredients that, along with implementation, ground an adequate account of physical computation are representation and mirroring.

Chapters 7 and 8 explain the semantic view of computation, fend off various objections, and lay out what Shagrir dubs the master argument for the semantic view (pp. 207ff). This argument is based on the idea that some physical systems can simultaneously implement different computational structures. Given the classification desideratum, there must be some constraint determining which computational structure is relevant, given both a certain context and a task to be performed by the system in that context. This constraint is semantic, namely, it is the semantic content of some of the representational states of the system. So, semantic properties are essential to individuating and classifying physical computing systems. But the semantic constraint, on its own, does not suffice to individuate physical computing systems. We also need some structure-preserving relationship to obtain between a putative computing system and its semantic properties, on the one hand, and the target domain where the system operates, on the other.

Chapter 9—which I found the most interesting—unpacks this last constraint, putting forward a view of computing as modelling. After introducing a relevant notion of modelling in the context of computation and illustrating it with a detailed case study concerning the neural integrator in the oculomotor system, Shagrir combines the three ingredients introduced in the previous chapters to define his modelling account of computation. According to this account, a physical system is a computing system just in case (a) the input–output function of a given process in the system 'mirrors' or preserves a certain relation (or structure) in a target domain such that this relation in the target domain and the input-output function characterizing the system's physical processes share some formal relation, *f*, (b) the physical process implements some formalism whose input–output function is *f*, and (c) the input and output variables of the physical process represent certain objects or properties in the target domain (p. 240).

Shagrir helpfully explains his account, referring to wider debates about the nature and roles of representation and computational modelling in the cognitive neurosciences (see, for example, Sprevak and Colombo [2018]). He discusses in some detail how modelling can help cognitive neuroscientists discover what function a given physical system computes, and explain why and how the system computes this function in the context of a given informationprocessing task. The upshot is a scientifically informed definition of a physical computing system that—although not obviously objective, even in the partial senses of objectivity Shagrir highlights in Chapter 1—is certainly not subjective either. This account allows us to distinguish computing from non-computing systems, to taxonomize distinct kinds of physical computing system, and, importantly, to do justice to the tremendous utility of computational modelling in the contemporary sciences of mind and brain.

Given its attention to salient features of actual scientific practice, meticulous reconstruction of philosophical arguments, and jargon-free writing style, *The Nature of Physical Computation* should be of interest to anybody who wants to make better sense of puzzling debates like that involving LaMDA and, more generally, anybody who wants to put into clearer focus some of the key concepts in the foundations of contemporary computational approaches to mind, brain, and behaviour.

Acknowledgements

Thanks to Nir Fresco for helpful comments on an earlier draft of this review.

Matteo Colombo
Tilburg Center for Logic, Ethics, and Philosophy of Science
m.colombo@uvt.nl

References

Chalmers, D. J. [2011]: 'A Computational Foundation for the Study of Cognition', *Journal of Cognitive Science*, **12**, pp. 325–59.

Colombo, M. [2021]: '(Mis)Computation in Computational Psychiatry', in F. Calzavarini and M. Viola (*eds*), *New Challenges in Philosophy of Neuroscience*, Cham: Springer, pp. 427–48.

Cummins, R. C. [1989]: *Meaning and Mental Representation*, Cambridge, MA: MIT Press.

Piccinini, G. [2020]: *Neurocognitive Mechanisms: Explaining Biological Cognition*, Oxford: Oxford University Press.

Schweizer, P. [2019]: 'Computation in Physical Systems: A Normative Mapping Account', in D. Berkich and M. V. d'Alfonso (*eds*), *On the Cognitive, Ethical, and Scientific Dimensions of Artificial Intelligence: Themes from IACAP 2016*, Berlin: Springer, pp. 27–47.

Sprevak, M. and Colombo, M. [2019]: *The Routledge Handbook of the Computational Mind*, London: Routledge.