BRAIN
BOOK REVIEW

Computational neuroscience: a grand unifying theory?

Artificial neural networks behave nothing like neurons in the brain. They have different capabilities, organization and dynamics. How can we bridge this gap between neurons and neural networks? Starting from the biological need to compute, Edmund Rolls asks, in his new book, how cognitive functions can be accomplished using a menagerie of neural motifs, in the context of the biological brain.

Few names in neuroscience span as broadly as Professor Rolls. From olfaction and taste, reward and punishment, to hippocampal auto-association and vision; from primate electrophysiology to human lesions, he has written on them all. But what runs through the whole of this new book is Rolls' commitment to unified algorithmic understanding of all cognition. This is hard to get from reading his papers, which are generally tightly focused.

Just shy of 1000 pages, this is one of the largest single-author neuroscience tomes and the first complete attempt to summarize our current knowledge about computation in the brain, at a level a graduate can understand.

As is traditional for neuroscience textbooks, it proceeds step by step through brain areas, starting with sensory, then motor pathways, then memory and high-level functions. But unlike a neuroscience textbook, each section starts with the computation, the formalisms, and only then considers the biology that underpins them. And unlike a computational textbook, biology is central, including neuromodulation, cortical layers, and information flow between brain areas.

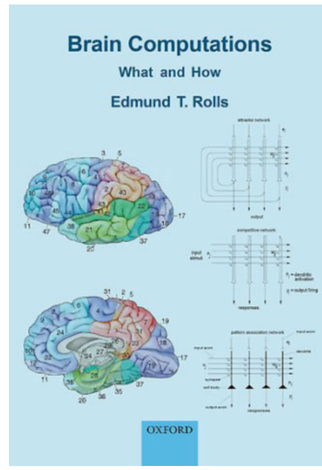
One major advantage of this innovative approach is that the models are supported not only by physiology, but by normative considerations. As a first question, Rolls asks, 'how *could* the task be accomplished algorithmically?', and only then, 'how can this work using biological neurons?' Rolls is true to David Marr.

The new and the old

Impressively recent findings are covered, including two-photon imaging, updates on functional MRI including hexadirectional coding, and modern reviews of language and semantics. At the same time, a wealth of critical historical studies from the 70s and 80s are summarized.

Some radical new explanations are proposed, for example how reverse replay might emerge from coupled timing circuits, how free will and creativity might arise from complexity and noise, and an innovative theory of how syntax could arise from role-specific 'word' neurons. On the other hand, the section on working memory mainly describes studies from 20 years ago. Higher level functions like cognitive control, belief and inference are less well covered, I suspect because we do not yet have the depth of evidence that Rolls would like.

Brain Computations: What and How
By *Edmund T. Rolls*, 2021
Oxford University Press
ISBN: 9780198871101
USD 80



Despite his deep physiological description, in most cases Rolls' model neurons remain simple one-compartment 'points'. Dendrite models are generally avoided. Computation is, at its heart, matrix multiplication. Simplification is the game—the maths is kept just complex enough to explain the phenomenon in question, yet intuitive enough to understand. Rolls seems to be telling us, you should be able to point your finger and say 'these axons signal *x*, these dendrites receive *y*'.

The unique 18th chapter presents an exposition of computational psychiatry. It compares psychiatry with fine-tuning a computational model. A programmer debugs an algorithm by observing how it misbehaves in certain situations; how it is not always robust. Similarly, a psychiatrist might consider the brain (e.g. in depression) to be less robust in some situations. The psychiatrist/programmer asks how we can tune the algorithm to function better in its environment. Moreover, we are led to the noble conjecture that experience in tuning neural networks should help in tuning mental health.

Astounding breadth

Brain Computations does not pretend to be a computational neuroscience textbook. Unlike textbook classics like Dayan and Abbott or Gerstner, which are for mathematicians admiring the biology, this book is for biologists admiring the maths. Rolls assumes some basic biological knowledge about neurons. Spiking and channel

Received December 13, 2021. Accepted December 13, 2021.

conductances are relegated to the appendix, and are apparently included to justify the more holistic approach adopted elsewhere. Instead, the book delves into the anatomy, neuronal organization and connectivity of the biological brain. Rolls' favoured scales and levels of explanation are cortical layers, subregions, mass models, selectivity, and synaptic learning rules like spike-timing dependent plasticity and synaptic normalization. This anchors the modelling to empirical work in a unique way. This is a biologically grounded, full systems neuroscience textbook—which makes it one of a kind.

The book links in ideas from deep neural networks, such as deep learning and backpropagation through time, but do not expect detailed coverage of the theory. Similarly, many explanations are framed using attractor dynamics, though dynamic systems theory is not itself covered. Major theoretical topics include sparseness, information content, associative learning and reward. To inform the arguments, data are mustered from functional imaging, lesion studies, single neuron work and comparative anatomy.

The book's ingredients certainly reflect Rolls' own palate. Hippocampus and orbitofrontal cortex are covered in more detail here than in any other computational book. Eight pages are devoted to hearing, compared to 25 on taste and 15 on olfaction. Reinforcement learning is relegated to the appendix apart from a couple of pages on the cingulate cortex. Model-based learning does not appear, and the topic of planning is given just six lines. Biologically, the thalamus gets four lines, and oscillations are mentioned only on a single page with an elegant but brutal take-down of the 'communication through coherence' hypothesis. By its balance of topics, the book carries the flavour of an unofficial autobiography.

Understanding by simulating

Rolls is a simulator: the equations read like computer code, and the book is accompanied by six MATLAB® code packages, complete with comments, some of which are also available in Python, designed to be a starting point for experimenting with the networks.

Although the detail makes many parts of brain function transparent, there are some sections that are more speculative and not formally captured in the models. For example, Chapter 9 presents beautifully crystallized simulations of hippocampal retrieval, but only vaguely sketches the route by which retrieval might influence ongoing cortical processing, resulting in a less testable and less predictive theory. Similarly, a rich theory of cingulate cortex is sketched, but a quantitative treatment that explains the data is conspicuously absent. While this is par for the course in neuroscience—nobody as yet has tied all of cognition together—the more speculative parts are not always clearly labelled, so readers do not get a firm grasp of where further work is warranted. This could have been an opportunity to highlight the dark corners that still need rigorous simulation, to complete the grand scheme. Rolls does underline language, especially syntax, as incomplete. But overall, he prefers to present things as 'solved', rather than highlighting the gaps.

A panoramic perspective

With so many parts of cognition modelled, how does one put them all together? One thing that is perhaps lacking is a final integrative stance. Can it all fit into a single framework? Presumably all this machinery is synthesized in the mind of Rolls, but for us mere readers, it is left to our imagination what it could be like to understand the whole brain.

Concealed in the book are several flash previews of the big picture—or at least of Rolls' genuine opinions. For example, an erudite attack is launched at large-scale networks identified using functional connectivity, which as Rolls highlights, miss the mark of explaining anything. The essential contribution of primate research is showcased. Unconventionally, we also get speculations on the cause of the world obesity epidemic, and a cameo mention of his recent theory of phenomenal consciousness.

Perhaps the book's biggest point is never made explicitly: 'A narrow portfolio of synaptic learning rules can buy you everything'. Hippocampal memories, action selection in the striatum, orbitofrontal reward representations, emotion in the limbic system, cerebellar motor control, parietal spatial coordinate transforms, place fields, and posterior visual object recognition—all these can emerge from relatively simple rules. This is Rolls' unspoken but substantial grand unifying theory.

Sanjay G. Manohar
Nuffield Department of Clinical Neurosciences, Level 6 West Wing, John Radcliffe Hospital, University of Oxford, Oxford, UK

E-mail: sanjay.manohar@ndcn.ox.ac.uk