The Computational Brain
by P. S. Churchland and T. J. Sejnowski

Bruce Bridgeman
Program in Experimental Psychology
University of California at Santa Cruz
Santa Cruz, Ca. 95064 USA

gbruce@file01.mpipf-muenchen.mpg.de

PSYCHE, 1(3), December 1993
http://psyche.cs.monash.edu.au/v1/psyche-1-3-bridgeman.html

Keywords: reductionism, neural networks, distributed coding, Karl Pribram, computational neuroscience, receptive field

1.1 The broad goal of this book, expressed at the start, is ``to understand how neurons give rise to a mental life." A mental reductionism is assumed in this seductively simple formulation. Indeed, the book represents reductionism at its best, as the authors guide the reader through the many intermediate levels that link neurons with mental life. In so doing they attack a problem that has persisted for some decades in the neurosciences, since the development of single-cell recording methods. The problem is that millions of neurons participate in every behaviorally meaningful activity, but we normally record from only one neuron at a time, or at best a handful. The temptation is great to overestimate the one-millionth sample obtained from a single neuron, to interpret its activity as detecting a perceptual situation or driving a motor response. This approach, seemingly inescapable in the 1960s, became untenable, but there were no concrete alternatives. Evoked potential techniques gave only a gross average of activity, too vague to pin down mechanisms, and early PDP (parallel distributed processing, or artificial neural network) models were too biologically unrealistic to provide viable interpretations of the single-cell data. Churchland and Sejnowski show how distributed models can now attack this problem, providing significant insights into brain function in a number of domains.

1.2 The book has several parts. First, the authors introduce their approach, combining anatomical, physiological, behavioral and modelling methods in an integrated interdisciplinary attack on specific functional systems. There follows a review of enough anatomy and neurophysiology to make the authors' viewpoint clear and to provide a background for integrating PDP modelling into specific problems in the neurosciences. The heart of the book is a series of chapters reviewing particular models that have been successful in increasing our understanding of the functioning of biological brains. Models of reflex reactions in invertebrates, of locomotion, the vestibulo-ocular reflex in primates, and several kinds of learning are analyzed. The authors wisely do a concentrated review of a few modelling efforts rather than attempting an encyclopedic coverage. Their examples successfully provide a flavor of what is possible and useful in the integration of brain modelling into the analysis of brain functions.

1.3 It is in putting brain models in perspective that the book is most useful, giving an informed and sensible view of how nervous systems are organized and how best to study them. Indeed, Sejnowski is successful in convincing even the skeptic of the usefulness of modelling distributed systems to understand neural organization. The approach echoes a modern conception of the brain not as the masterful, unified work of God, but as a multifaceted contraption pieced together through evolution, using whatever was at hand to solve problems and winding up with a collection of special-purpose systems that collectively do the job. Evolution is not engineering, a fact to be kept in mind when attempting the reverse engineering of computational neuroscience.

1.4 The ``bag of tricks'' concept of brain architecture is echoed in the new generation of distributed models. Gone are the abstract models that attempt to solve a complex problem in a single elegant layer. They have been replaced by homeomorphic models, in which every part of the model, even the distributed net, has a counterpart in known physiology or anatomy. A net is embedded in a physiologically realistic context, and sometimes several nets are modeled at the places where the neuroanatomy calls for them. Though the general architecture of all the net models is similar, the specifics are adapted to the problem at hand---should the weights (the strengths of connections among the modeled units) be set by hand, or by some self-instructing algorithm? Is recurrent feedback necessary? Should the inputs be filtered, transformed, selected? How should the model be restricted within the infinitely extendible domain of brain function? In such questions Sejnowski provides reasoned and reasonable arguments. He also realistically defines and analyzes the limitations of the models, what they cannot do as well as what they can.

1.5 A chief benefit of PDP models is the power of representing information in combinations of activities of many units. The basic idea is that a dozen detector cells considered separately can encode only 12 states, for example, while 12 binary units considered in combination can detect 2 to the 12th power or 4096 states. As the number of units escalates to brain proportions, the advantages of combinatoric coding become overwhelming. Though this idea has been around for a long time (i.e., Bridgeman, 1971), it was neglected in the enthusiasm for detectors of lines, edges, faces, and even hands. The book points out how vector coding can combine combinatoric efficiency with distributed processing to retrieve separate messages from overlapping sets of simulated neurons. The idea is not a new one; it was standard fare for Cornell undergraduates in the mid-1960s, and the mathematics is a first cousin of the orthogonal polynomials known to every student of statistics. But here distributed coding is combined with powerful simulation techniques and with architectures that make the ideas realizable.

1.6 The new approach distinguishes several types of distributed codes, with vector coding for instance being useful in some situations and vector averaging in others. The approach can make sense of the myriad of different neural responses, each neuron seemingly having its own ``personality'' of response profile, taking combinatoric advantage of the diversity to create distinct vector codes for every possible situation. The approach is causing a re-evaluation of data that had been uninterpretable under the old feature detector scheme, but fit well with vector coding (Bridgeman, 1992).

1.7 The models succeed in some surprising places. Several models replicate the experimentally observed properties and distributions of receptive fields or action profiles, the characteristics of the outside world to which the sensory neuron is sensitive. Biological and computer models converge on similar receptive field properties even though the receptive fields are created by the models themselves and not by the modeler. The models and the evolutionary process may find optima by very different methods, but the fact that the optima correspond even roughly is an impressive indication that the models are on the right track. Again, the wide variety of neural responses becomes a solution rather than a problem in this context. A striking example of this is in a model that finds shape from shading: the model develops receptive fields for some of its intermediate-layer units that are sensitive to edges, though the model was never exposed to edges. This success should also warn us of the danger of mistaking a receptive field's profile for its function.

1.8 There are some caveats. The authors note that they have carved out a large goal for a single volume, and some corners will have to be cut. They are not always cut in the right places, however. The neuron described in the introduction (p. 3), for example, is a simplistic one with a feedforward dendrite-to-axon architecture, and though the picture is later made more accurate with dendro-dendritic information flow and biochemical variety, the emphasis is still on a 1930s neuron.

1.9 Though the book is generally accurate, its creativity is a more important asset than its scholarship. The authors admit to a California bias, but sometimes their here-and-now approach neglects a history that can prevent mistakes from being repeated. A glaring omission, in a book that champions distributed coding, is the work of Karl Pribram, who for two decades at Stanford was practically the only leading proponent of such coding. Pribram (i.e., 1971) never accepted the detector approach to neural coding; in the 1960s and 1970s, when detector cell theories dominated, he stubbornly insisted that the neurological evidence for distributed coding could not be ignored. Another lapse of historical perspective is in characterizing the 1952 Hodgkin-Huxley equation for the nerve impulse as the first quantitative model in the neurosciences. The Nernst equation of a half-century earlier, however, quantitatively described the static voltage across the membrane from ion concentrations and chemical principles, and is the basis upon which the Hodgkin-Huxley model is built. And a century-long history of work on memory consolidation, with complex and contradictory conclusions, is neglected in favor of a single recent experiment.

1.10 The uncritical description of distributed models as connectionist also neglects a history, this time one that reversed the meaning of the term. Connectionistic models originally were those of Pavlov and the subsequent behaviorists, who explained links between perceptions and motor responses with anatomical connections between sensory and motor centers in the brain. A behavioral link between perception A and action B implied activation of axonal connections from A to B. Physiological work failed to find such connections. As explicit alternatives to the old connectionist concepts, the first PDP models were called ``neo-connectionist." Before long the ``neo" was dropped, and the term came to mean the exact reverse of its original reference. So at present, the term connectionist really means non-connectionist.

1.11 There is redundant organization in several parts of the book. The basic PDP architecture is described twice, once near the start and again in a chapter on perceptual processes, and even an example of generalization using the word ``red" is given twice. And there are some mistakes, especially in the neurophysiology which is the speciality of neither author. Muscle organization is described as push-pull (p. 361), for example, when it is the control system that is so organized. The muscles can only pull. Graded potentials, where the real combining of information takes place on dendrites and cell somas, receive scant notice and are not mentioned by name. Dominant in the text are the action potentials, which merely transfer information from one place to another without changing it. Visual receptive fields in the cortex are interpreted without benefit of the modern spatial-frequency analysis that lends itself to mathematical modelling, and describes cell responses better than the old line and edge metaphors. The distinction is now common even at the textbook level (Bridgeman, 1988).

1.12 The writing style varies somewhat through the text. While Sejnowski is clear and down-to-earth, Churchland possesses a marvellously rich and evocative style, flush with original metaphors and clever devices. It sometimes borders on the self-consciously cute, as in the description of three kinds of long-term potentiation as vanilla, chocolate, and strawberry ripple; but the writing is generally effective.

1.13 Notwithstanding these relatively minor quibbles, the volume stands a a significant achievement in bringing the technically challenging world of distributed modelling to the non-specialist reader, and in intelligently integrating PDP modelling into neurophysiology. Computational modelling has joined ``wet" biological techniques in the essential armamentarium of modern brain science.

References

Bridgeman, B. (1971). Metacontrast and lateral inhibition. *Psychological Review, 78*, 528-539.

Bridgeman, B. (1988). *The biology of behavior and mind*. New York: Wiley.

Bridgeman, B. (1992). Taking distributed coding seriously. *Behavioral and Brain Sciences, 15*, 717.

Pribram, K. H. (1971). *Languages of the brain*. Englewood Cliffs, N. J.: Prentice-Hall.