From the Retina to the Neocortex

Selected Papers of David Marr

Edited by Lucia Vaina

With Commentaries by

Jack D. Cowan
W. Eric L. Grimson
Norberto M. Grzywacz
Ellen C. Hildreth
Bruce McNaughton
Terrence J. Sejnowski
W. Thomas Thach
David Willshaw

Birkhäuser
Boston · Basel · Berlin 1991

David Marr: A Pioneer in Computational Neuroscience

Terrence J. Sejnowski

David Marr: A Pioneer in Computational Neuroscience

David Marr advocated and exemplified an approach to brain modeling that is based on computational sophistication together with a thorough knowledge of the biological facts. The pioneering papers in this collection demonstrate that a combination of computational analysis and biological constraints can lead to interesting neural algorithms. The recent developments in computational models of neural information processing systems is an extension of this seminal research: Marr has influenced the latest generation of network models through both his models and his emphasis on the computational level of analysis (Marr, 1975, 1982). Progress has been made by adopting an integrated approach in which constraints from all three of Marr's levels of analysis—the computational, algorithmic and the implementational—are applied at many different levels of investigation (Sejnowski and Churchland, 1989).

These early papers are not easy to read. Marr gives the reader too many concrete details and too little overall guidance. He demands of the reader a deep understanding of probability theory and an encyclopedic knowledge of neuroanatomy. Even those who are steeped in the current generation of neural network modeling will find terms in the early papers difficult to translate into recent usage. Still, they are reminiscent of the style found in Maxwell's papers, which were written before the invention of vector notation. Just as Maxwell introduced specialized terminology and drew analogies with mechanical concepts such as gears and idler wheels, there are unusual terms in Marr's papers, such as "codons," borrowed from molecular genetics, and a novel set of concepts that must be mastered before the papers can be appreciated. Although some of the central ideas in these early papers are well known, there are important insights into neural computation that will handsomely repay the reader the effort taken to master the terminology.

The first four papers in this collection are from Marr's Cambridge period in the 1960s, among the earliest in his career, and articulate a remarkably ambitious theory of memory. These models were firmly based on what was then known about the structure of the brain. Even those who know that Marr had a model for motor learning in the cerebellum (Marr, 1969) may not be aware that his models of neocortex (Marr, 1970) and the hippocampus (Marr, 1971) were even more detailed than his cerebellar model. In a personal conversation with me in 1975, Marr singled out the neocortex paper as the one among these early papers that he was most proud of. The last set of papers on vision in this collection, are transitional and reflect a new, computationally-motivated approach to vision that guided his later research at MIT. At the same time they reflect an evident fascination with the structure of the brain, a fascination that Marr retained throughout his career.

The learning algorithm in the cerebellar model requires an external "teacher" to instruct the synapses and would today be called a supervised learning procedure. This type of error-correction learning has been extended to multiple layers of processing units (Rumelhart et al., 1986). The archicortex model uses a form of unsupervised learning that is based on competition among the output units. Similar unsupervised learning algorithms for neural network models have recently been studied for clustering data (Grossberg, 1976; Kohonen, 1984; Rumelhart and Zipser, 1985). They are "simple" memory systems that do not address the central issue of how the information is represented on the inputs and outputs.

In contrast to the simple memory models, the neocortical model is about category formation and discovery of high-order patterns in data based on unsupervised learning. Several recent models have been proposed to attack the problem of extracting information from sensory data (Hinton and Becker, 1990; Linsker, 1990). Marr's approach was different and depended on recruiting new neurons to represent high-order statistical structure or "concepts." Unfortunately, the computational resources available to Marr in the late 1960s were minimal, and one of the frustrations when reading these papers is the lack of numerical simulations. Promising algorithms do not always perform as expected when confronted with data from the real world; too many simplifications must be made so that the analysis is tractable. Until simulations of the neocortex model are performed we will not be able to assess its effectiveness.

The retina paper is an attempt to match a computational problem—finding the lightness of a surface—to anatomical substrates in the retina (Marr, 1974). This paper is transitional: Marr's subsequent papers in vision would emphasize more and more the computational aspects while relying less and less on anatomy. The predictions in this paper were not borne out by subsequent and more recent physiological recordings from single neurons make it likely that the locus of lightness and color constancy is in visual cortex (Zeki, 1983). Interestingly, recent algorithms based on neural network models are similar to those proposed by Marr (Hurlbert and Poggio, 1988; Land, 1986). Old ideas often come back in contexts that their originators might not even recognize.

Perhaps the best known papers in this collection are the models of binocular depth perception (Marr and Poggio, 1976, 1979). These models were appealing because they were based on plausible biological mechanisms, were constrained by psychophysical data and were tested by simulations on real-world data. In the first model, Marr and Poggio performed simulations to demonstrate the effectiveness of their algorithm on random-dot stereograms, first introduced by Bela Julesz (Julesz, 1971). The elegant simplicity of the network model was anticipated in earlier research (Dev, 1975; Nelson, 1975), but the interpretation of the constraints and the convincing demonstration made this a landmark paper. This approach to constraint satisfaction was an inspiration for connectionist-style models of visual computation (Ballard et al. 1983).

One of the remarkable properties of the Marr-Poggio stereo network was that it always converged. The stereo network is a highly nonlinear system of equations and attempts to analyze the nonlinear equations came to the conclusion that they were as difficult to analyze as Conway's game of "life" (Marr et al. 1978). In 1982, John Hopfield pointed out that symmetric networks like the Marr-Poggio model were a special case because they possess an energy or Lyapunov function that guarantees convergence to a local energy minimum (Hopfield, 1982). The stereo network was designed in such a way that the local energy minima are solutions to the problem. It is also possible to design network models to handle transparent surfaces in random-dot stereograms (Qian and Sejnowski, 1988). Subsequent developments showed how even more difficult constraint satisfaction problems can be solved by globally minimizing the energy (Hopfield and Tank, 1986; Kienker et al., 1986; Poggio et al., 1988). Marr later felt that the time delays inherent in the relaxation of a network to a solution were unsatisfactory given the speed with which our visual system can interpret most images (Marr, 1982: see p. 107). The shortcomings of the first stereo model were addressed in the second model (Grimson, 1981; Marr and Poggio, 1979), which was much faster and took into account multi-resolution filters that could be applied to real images. However, there are other aspects of stereo vision that cannot be handled by this algorithm. The human visual system is even more clever than these early stereo algorithms (Poggio and Poggio, 1984).

David Marr continued to make major contributions to the study of vision. Those who have been influenced primarily by Marr's book on the computational approach to vision (Marr, 1982) may be surprised by the extraordinary attention given in this collection of papers to neuroanatomy. In rereading them, it is possible to put into perspective Marr's later work on the computational approach to vision. Although Marr made fewer appeals to detailed biological mechanisms in his later vision papers, there were still many examples of inspiration and confirmation of computational approaches from neuropsychological and psychophysical data and constraints from physiological measurements. The scientific style of the papers in this collection make it clear that these intrusions from the biological realm were not incidental.

One of the inevitable problems of building models and theories in neuroscience is that new facts about the brain are continually being discovered and old ideas are sometimes modified or discarded. The striking advances in neuroscience since these early papers are most evident in our present view of neurons. In 1970, dendrites were thought to be passive cables and ideas on synaptic mechanisms were based primarily on the neuromuscular junction. Today, dendrites are known to have voltage-dependent conductances that make them dynamical entities (Llinas, 1988); a gallery of channels and neurotransmitter receptors with a wide range of time scales allow neurons to burst and oscillate, and synapses to potentiate and habituate (Kandel et al., 1987). Marr's models need to be updated to take these new properties into account. However, the insight that a powerful computational system could be built from a sophisticated model of memory remains an exciting idea, and the goal of incorporating anatomical constraints into network models of vision is now being actively pursued (Sejnowski et al., 1988).

Finally, how is one to reconcile the research direction implicit in these papers and the explicit statements found in Marr (1982) regarding the independence of the computational level from the implementation level? This principle, taken out of the context provided by Marr's research style, gives the misleading impression that constraints from the algorithmic and implementation levels found in biological systems are unnecessary. A remarkable feature of Marr's book is the degree to which biological considerations enter on almost every page in inspiring computational analysis, in choosing between algorithms and in providing the ultimate measure of success. Computational explanations for our visual and mental abilities eventually may be found, and seeking such explanations is essential—this was Marr's message. However, he was far from abandoning biological and psychological data in reaching this goal.

The performance of our perceptual and cognitive systems and the way that brains are organized provide essential constraints on possible computational explanations (Churchland and Sejnowski, 1988; Sejnowski et al., 1989). Neural circuits and how they function clearly inspired Marr and they continue to be rich sources of inspiration for many of us.

REFERENCES

Ballard D, Hinton G, Sejnowski T (1983): Parallel visual computation. *Nature* 306:21-26

Churchland PS, Sejnowski TJ (1988): Perspectives on cognitive neuroscience. *Science*, 242:741-745

Dev P (1975): Perception of depth surfaces in random-dot stereograms: a neural model. *Int J Man-Machine Stud* 7:511-528

Grimson E (1981): *From Images to Surfaces*. Cambridge, MA: MIT Press

Grossberg S (1976): Adaptive pattern classification and universal recoding: I: Parallel development and coding of neural feature detectors. *Biol Cybern* 23: 121-134

Hinton GE, Becker S (1990): *An unsupervised learning procedure that discovers surfaces in random-dot stereograms*. New Jersey: Lawrence Erlbaum Associates, pp 218-222

Hopfield JJ, (1982): Neural networks and physical systems with emergent collective computational abilities. *Natl Acad Sci USA* 79: 2554-2558

Hopfield JJ, Tank DW (1986): Computing with neural circuits: A model. *Science* 233: 625-633

Hurlbert AC, Poggio TA (1988): Synthesizing a color algorithm from examples. *Science* 239: 482-485

Julesz B (1971): *Foundations of Cyclopean Vision*. Chicago: University of Chicago Press

Kandel ER, Klein M, Hochner B, Shuster M, Siegelbaum SA, Hawkins RD, Glanzman DL, Castellucci VF (1987): Synaptic modulation and learning: new insights into synaptic transmission from the study of behavior. In *Synaptic Function*, Edelman GM, Gall WE, Cowan WM, eds, New York: John Wiley and Sons, pp 471-518

Kienker PK, Sejnowski TJ, Hinton GE, Schumacher LE (1986): Separating figure from ground with a parallel network. *Perception* 15:197-216

Kohonen T (1984): *Self-Organization and Associative Memory.* New York: Springer Verlag

Land EH (1986): An alternative technique for the computation of the designator in the retinex theory of color vision. *Proc Natl Acad Sci USA* 83:3078-3080

Linsker R (1990): Perceptual neural organization: some approaches based on network models and information theory. *Annu Rev Neurosci* 13:257-281

Llinas RR (1988): The intrinsic electrophysiological properties of mammalian neurons: Insights into central nervous system function. *Science* 242:1654-1664

Marr D (1969): A theory of cerebellar cortex. *J Physiol Lond* 202:437-470

Marr D (1970): A theory for cerebral neocortex. *Proc R Soc Lond, B* 176:161-234

Marr D (1971): Simple memory: a theory for archicortex. *Phil Trans Roy Soc B* 262: 23-81

Marr D (1974): The computation of lightness by the primate retina. *Science* 14:1377-1387

Marr D (1975): Approaches to biological information processing. *Science* 190: 875-876

Marr D (1982): *Vision.* San Francisco: Freeman

Marr D, Palm G, Poggio T (1978): Analysis of a cooperative stereo algorithm. *Biol Cybern* 28: 223-239

Marr D, Poggio T (1976): Cooperative computation of stereo disparity. *Science* 194::283-287

Marr D, Poggio T (1979): A computational theory of human stereo vision. *Proc R Soc Lond Ser. B* 204: 301-28

Nelson, JI (1975): Globality and stereoscopic fusion in binocular vision. *J Theor Biology* 49:1-88

Poggio G, Poggio T (1984): The analysis of stereopsis. *Annu Revs Neurosci* 7:379-412

Poggio T, Gamble EB, Little JJ (1988): Parallel integration of vision modules. *Nature* 242:436-439

Qian N, Sejnowski TJ (1988): *Learning to solve random-dot stereograms of dense transparent surfaces with recurrent backpropagation.* Pittsburgh, PA: Morgan-Kaufmann Publishers, pp 235-443

Rumelhart D, Zipser D (1985): Feature discovery by competitive learning. *Cogn Sci* 9:75-112

Rumelhart DE, Hinton GE, Williams RJ (1986): Learning internal representations by error propagation. In: *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 1: Foundations* Cambridge, MA: MIT Press

Sejnowski T, Koch C, Churchland P (1988): Computational neuroscience. *Science* 241:1299-1306

Sejnowski TJ, Churchland PS (1989): Brain and cognition. In: *Foundations of Cognitive Science,* Posner MJ, ed. Cambridge, MA: MIT Press

Zeki S (1983): Colour coding in the cerebral cortex: the reaction of cells in monkey visual cortex to wavelengths and colors. *Neuroscience* 9:741-765

Professor
The Salk Institute
and the University of California, San Diego
La Jolla, California