David Marr

Shimon Edelman

Department of Psychology, 232 Uris Hall

Cornell University, Ithaca, NY 14853-7601, USA

Lucia M. Vaina

Department of Biomedical Engineering and Neurology

Boston University, 44 Cummington St., Boston, MA 02215, USA

David Courtnay Marr was born on January 19, 1945 in Essex, England. He attended Rugby, the English public school, on a scholarship, and went on to Trinity College, Cambridge. By 1966, he obtained his B.S. and M.S. in mathematics, and proceeded to work on his doctorate in theoretical neuroscience, under the supervision of Giles Brindley. Having studied the literature for a year, Marr commenced writing his dissertation. The results, published in the form of three journal papers between 1969 and 1971, amounted to a theory of mammalian brain function, parts of which remain relevant to the present day, despite vast advances in neurobiology in the past three decades. Marr's theory was formulated in rigorous terms, yet was sufficiently concrete to be examined in view of the then available anatomical and physiological data. Between 1971 and 1972, Marr's attention shifted from general theory of the brain to the study of vision. In 1973, he joined the Artificial Intelligence Laboratory at the Massachusetts Institute of Technology as a visiting scientist, taking on a faculty appointment in the Department of Psychology in 1977, where he was made a tenured

full professor in 1980. In the winter of 1978 he was diagnosed with acute leukemia. David Marr died on November 17, 1980, in Cambridge, Massachusetts. His highly influential book, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, which has redefined and revitalized the study of human and machine vision, has been published posthumously, in 1982.

Marr's initial work in neuroscience combined high-level theoretical speculation with meticulous synthesis of the anatomical data available at the time. The question he chose to address is the *nec plus ultra* of neuroscience: what is it that the brain does? Marr proposed a definite answer to this question for each of three major brain structures: archicortex (the phylogenetically older part of the cerebral cortex), cerebellum, and neocortex. The three answers complement each other, rallying around the idea that the brain's central function is statistical pattern recognition and association, carried out in a very high-dimensional space of "elemental" features. The basic building block of all three theories is a *codon*, or a subset of features, with which there may be associated a cell, wired so as to fire in the presence of that particular codon.

In the first paper, Marr proposed that the cerebellum's task is to learn the motor skills involved in performing actions and maintaining posture (Marr, 1969). The Purkinje cells in the cerebellar cortex, presumably implementing the codon representation, associate (through synaptic modification) a particular action with the context in which it is performed. Subsequently, the context alone causes the Purkinje cell to fire, which in turn precipitates the next elemental movement. Thirty years later, a significant proportion of researchers working on the cerebellum seem to consider this model as "generally correct" — a striking exception in a field where the *nihil nisi bene* maxim is not known to be observed.

The next paper (Marr, 1970) extended the codon theory to encompass a more general kind of

statistical concept learning, which he assessed as "capable of serving many of the aspects of the brain's function" (the vagueness of this aspect of the theory would lead him soon to abandon this approach, which, as he realized all along, was "once removed from the description of any task the cerebrum might perform"). How can a mere handful of techniques for organizing information (such as the codon representation) support a general theory of the brain function? Marr's views in this matter are profoundly Realist, and are based on a postulate of "the prevalence in the world of a particular kind of statistical redundancy, which is characterized by a 'Fundamental Hypothesis'," stating that "Where instances of a particular collection of intrinsic properties (i.e., properties already diagnosed from sensory information) tend to be grouped such that if some are present, most are, then other useful properties are likely to exist which generalize over such instances. Further, properties often are grouped in this way" (Marr 1970, pp. 150-151). These ideas presaged much of the later work by others on neural network models of brain function, which invoke the intuition of learning as optimization ("mountain climbing") in an underlying probabilistic representation space.

A model at whose core is the tallying of probabilities of events needs an extensive memory of a special kind, allowing retrieval based on the content, rather than the location, of the items. Marr's third theoretical paper considers the hippocampus as a candidate for fulfilling this function (Marr, 1971). In analyzing the memory capacity and the recall characteristics of the hippocampus, Marr integrated abstract mathematical (combinatorial) constraints on the representational capabilities of codons with concrete data derived from the latest anatomical and electrophysiological studies. The paper postulated the involvement in learning of synaptic connections modifiable by experience — a notion originating with the work of Donald Hebb in the late 1940's, and discussed by Marr's mentor Brindley in a 1969 paper. Marr provided a mathematical proof of efficient partial content-based recall by his model, and offered a functional interpretation of many anatomical structures in the

hippocampus, along with concrete testable predictions. Many of these (such as the existence in the hippocampus of experience-modifiable synapses) were subsequently corroborated (see the reviews in (Vaina, 1990)).

A consummation of this three-pronged effort to develop an integrated mathematical-neurobiological understanding of the brain would in any case have earned Marr a prominent place in a gallery, spanning two and a half centuries (from John Locke to Kenneth Craik), of British Empiricism, the epistemological stance invariably most popular among neuroscientists. As it were, having abandoned the high-theory road soon after the publication of the hippocampus paper, Marr went on to make his major contribution to the understanding of the brain by essentially inventing a field and a mode of study: *computational neuroscience*. By 1972, the focus of his thinking in theoretical neurobiology shifted away from abstract theories of entire brain systems, following a realization that without an understanding of specific tasks and mechanisms — the issues from which his earlier theories were "once removed" — any general theory would be glaringly incomplete.

Marr first expressed these views in public at an informal workshop on brain theory, organized in 1972 at the Boston University by Benjamin Kaminer. In his opening remarks, he suggested an "inverse square law" for theoretical research, according to which the value of a study varies inversely with the square of its generality — an assessment that favors top-down reasoning anchored in functional (computational) understanding, along with bottom-up work grounded in an understanding of the mechanism, but not theories derived from intuition, or models built on second-hand data.

The new methodological stance developed by Marr following the shift in his views is summarized in a remarkably lucid and concise form in a two-page book review in *Science*, titled "Approaches to Biological Information Processing" (Marr, 1975). By that time, Marr came to believe firmly that the field of biological information processing had not yet accrued an empirical basis sufficient for

guiding and supporting a principled search for a general theory. Remarking that the brain may turn out to admit "of no general theories except ones so unspecific as to have only descriptive and not predictive powers" — a concern echoed in one of his last papers (Marr, 1981) — he proceeded to mount a formidable critique of the most common of the theories circulated in the early 1970s, such as catastrophe theory and neural nets (the current popularity of dynamical systems and of connectionism, taken along with the integration of Marr's critical views into the mainstream theoretical neurobiology, should fascinate any student of the history of ideas).

The main grounds for his argument, which was further shaped by an intensive and fruitful interaction with Tomaso Poggio (Marr and Poggio, 1977), were provided by an observation that subsequently grew into a central legacy of Marr's career: the understanding of any information processing system is incomplete without insight into the problems it faces, and without a notion of the form that possible solutions to these problems can take. Marr and Poggio termed these two levels of understanding *computational* and *algorithmic*, placing them above the third, *implementational*, level, which, in the study of the brain, refers to the neuroanatomy and neurophysiology of the mechanisms of perception, cognition, and action.

Upon joining the MIT AI Lab, Marr embarked on a vigorous research program seeking computational insights into the working of the visual system, and putting them to the test of implementation as computer models. Marr's thinking in the transitional stage, at which he treated computational results on par with neurobiological findings, is exemplified by the paper on the estimation of lightness in the primate retina (Marr, 1974); subsequently, much more weight was given in his work to top-down, computational-theory considerations. This last period in Marr's work is epitomized by the theory of binocular stereopsis, developed in collaboration with Poggio, and presented in a series of ground-breaking papers (Marr and Poggio, 1976; Marr and Poggio, 1979). At that time,

Marr also worked on low-level image representation (Marr, 1976; Marr and Hildreth, 1980), and on shape and action categorization (Marr and Nishihara, 1978; Marr and Vaina, 1982). Marr's book, *Vision*, written during the last months of his life, is as much a summary of the views of what came to be known as the MIT school of computational neuroscience as it is a personal credo and a list of achievements of the second part of Marr's scientific endeavor, which lasted from about 1972 to 1980.

The blend of insight, mathematical rigor, and deep knowledge of neurobiology that characterizes Marr's work is reminiscent of the style of such titans of neuroscience as Warren McCulloch — except that McCulloch's most lasting results were produced in collaboration with a mathematician (Walter Pitts), whereas Marr did his own mathematics. A decade after his quest was cut short, it has been claimed both that Marr is cited more than he is understood (Willshaw and Buckingham, 1990), and that his influence permeates theoretical neurobiology more than what one would guess from counting citations (McNaughton, 1990). Still, contributors to the mainstream journals in neurobiology now routinely refer to the "computations" carried out by the brain, and the most exciting developments are those prompted (or at least accompanied) by computational theories.

In computer vision (a branch of artificial intelligence), the influence of Marr's ideas has been complicated by the dominance of the top-down interpretation of his methodology: proceeding from a notion of what needs to be done towards the possible solutions. For some time, Marr's school was identified there with the adherents of a particular computational theory of vision, which claims that constructing an internal model of the world is a prerequisite for carrying out any visual task. The accumulation of findings to the contrary in neurobiology and in the behavioral sciences gradually brought to the fore the possibility that vision does not require geometric reconstruction. This encouraged researchers to seek alternative theories, some of which employ concepts and techniques

that did not exist in the 1970s, or were not known to the scholars of vision at the time. These new ideas, in turn, are making their way into neuroscience, as envisaged by Marr.

On a more general level, Marr's work provided a solid proof that a good theory in behavior and brain sciences need not have to trade off mathematical rigor for faithfulness to specific findings. More importantly, it emphasized the role of explanation over and above mere curve fitting, making it legitimate to ask *why* a particular brain process is taking place, and not merely what differential equation can describe it.

References

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

Marr, D. (1970). A theory for cerebral neocortex. *Proceedings of the Royal Society of London B*, 176:161–234.

Marr, D. (1971). Simple memory: a theory for archicortex. *Phil. Trans. Royal Soc. London*, 262:23–81.

Marr, D. (1974). The computation of lightness by the primate retina. *Vision Research*, 14:1377–1388.

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

Marr, D. (1976). Early processing of visual information. *Phil. Trans. R. Soc. Lond. B*, 275:483–524.

Marr, D. (1981). Artificial intelligence: a personal view. In Haugeland, J., editor, *Mind Design*, chapter 4, pages 129–142. MIT Press, Cambridge, MA.

Marr, D. and Hildreth, E. (1980). Theory of edge detection. *Proc. R. Soc. Lond. B*, 207:187–217.

Marr, D. and Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three dimensional structure. *Proceedings of the Royal Society of London B*, 200:269–294.

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

Marr, D. and Poggio, T. (1977). From understanding computation to understanding neural circuitry. *Neurosciences Res. Prog. Bull.*, 15:470–488.

Marr, D. and Poggio, T. (1979). A computational theory of human stereo vision. *Proceedings of the Royal Society of London B*, 204:301–328.

Marr, D. and Vaina, L. (1982). Representation and recognition of the movements of shapes. *Proceedings of the Royal Society of London B*, 214:501–524.

McNaughton, B. L. (1990). Commentary on Simple memory: a theory of the archicortex. In Vaina, L. M., editor, *From the retina to the neocortex: selected papers of David Marr*, pages 121–128. Birkhauser, Boston, MA.

Vaina, L. M., editor (1990). *From the retina to the neocortex: selected papers of David Marr.* Birkhauser, Boston, MA.

Willshaw, D. J. and Buckingham, J. T. (1990). An assessment of Marr's theory of the hippocampus as a temporary memory store. *Proceedings of the Royal Society of London B*, 329:205–215.