

Marr, David

Lucia M Vaina, Boston University and Harvard Medical School, Boston, Massachusetts, USA

CONTENTS

Introduction

Contributions to theoretical neuroscience: what is it that the brain does?

A pioneer of computational neuroscience

A computational theory of vision: how does the brain see?

David Marr's vision

INTRODUCTION

0456.001 David Courtnay Marr was born on 19 January 1945 in Essex, England. He went to the English public school, Rugby, on a scholarship, and between 1963 and 1966 he studied at Trinity College, Cambridge, where he obtained his BA degree in mathematics with first-class honors. For his doctoral research he continued in theoretical neuroscience, under the supervision of Giles Brindley. His education involved training in neuroanatomy, neurophysiology, biochemistry and molecular biology. At Trinity College in 1971 he received an MA (with distinction) in mathematics and a PhD in theoretical neurophysiology. After obtaining his PhD, he accepted a research appointment at the Medical Research Council (MRC) Laboratory of Molecular Biology under Sydney Brenner and Francis Crick, and he retained an affiliation with the MRC until 1976.

CONTRIBUTIONS TO THEORETICAL NEUROSCIENCE: WHAT IS IT THAT THE BRAIN DOES?

0456.002 In three successive papers that combine high-level theoretical speculation with meticulous synthesis of the available neuroanatomical data, Marr proposed a definite answer to this question for the cerebellum, archicortex and neocortex. Common to these three studies is the idea that the central function of the brain is statistical pattern recognition and association in a very high-dimensional space of 'elemental' features. The basic building block of all three theories is the codon, or a subset of features, with which a cell that is wired in such a way as to fire in the presence of that particular codon is associated.

0456.003 A paper entitled 'A theory of cerebellar cortex,' published in 1969 (see Further Reading), represents

the essence of Marr's doctoral research. The paper is a theoretical model that made critical predictions elucidating how the cerebellum learns the motor skills involved in performing actions and maintaining posture and balance. The fundamental elements of the model are the known cell types in the cerebellum, their connectivities and the synaptic actions of the cerebellar cortex. The process involves context recognition and learning. The former was described at the level of the mossy fiber-granule cell-Golgi cell circuitry, and the latter was described at the level of the parallel fiber-Purkinje cell synapse, heterosynaptically strengthened by the inferior olive climbing fiber. Linked through learning to the context of the previous movement in the sequence, the Purkinje cell, presumably implementing the codon representation, associates (through synaptic modification) a particular movement 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. Basically the cerebellum model is a one-layer network (of granule cells) with fixed synapses, and an associative memory store and a set of conditioning inputs (the climbing fibers).

The second paper, entitled 'A theory for cerebral neocortex', published by Marr in 1970 (see Further Reading), extended the codon theory to encompass a more general type of statistical concept learning, which he assessed as being 'capable of serving many of the aspects of the brain's functions', in particular the formation and organization of networks capable of classifying and representing 'the world'. This hypothesis is an early attempt at a theory of unsupervised learning relating to methods of cluster analysis. The paper discusses the structure of the relationships which appear in the afferent information, and the usefulness to the organism of discovering them. These two ideas are

0456.004

combined by the ‘fundamental hypothesis’ which is based on the existence and prevalence in the world of a particular type of ‘statistical redundancy’. The fundamental hypothesis, as set out in Marr’s 1970 paper, states 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.

0456.005

The neocortex model keeps track of probabilities of events, and to do this it needs an extensive memory of a special kind, allowing retrieval that is based on the content rather than the location of the items. In his third theoretical paper, entitled ‘Simple memory: a Theory for Archicortex’, published in 1971 (see Further Reading), Marr considers the hippocampus as a candidate for fulfilling this function. In analyzing the memory capacity and recall characteristics of the hippocampus, Marr integrated combinatorial–mathematical constraints on the representational capabilities of codons with concrete data derived from neuroanatomical and neurophysiological studies. In modern terms, the hippocampal model consists of a recurrent network with two layers of trainable ‘hidden’ units that encode and classify input patterns connected to an associative memory store. The paper postulated the involvement in learning of synaptic connections modifiable by experience – a notion that originated from the research of Donald Hebb in the late 1940s. The paper is a mathematical proof of efficient partial content-based recall by the model, and it offered a functional interpretation of many anatomical structures in the hippocampus, together with concrete testable predictions.

0456.006

‘Truth, I believed, was basically neuronal, and the central aim of research was a thorough analysis of the structure of the nervous system’. This view expressed by Marr in 1982 in his book, *Vision: a Computational Investigation into the Human Representation and Processing of Visual Information* (see Further Reading), combined with his initial training in mathematics, shaped the quantitative, analytical methodology that he applied in these three studies. In a letter to Francis Crick in 1977 he summarizes his fundamental views as follows:

For a mathematician, understanding (or explanation) is all, yet in science proof is, of course, what counts. In the case of information-processing devices, understanding is very important; one can know a fact about a device for years without really understanding it, and

part of the theoretician’s job is to place into a comprehensible framework the facts that one already knows. I still think that the cerebellum is a good example. For sure, the idea that the parallel fiber–Purkinje cell synapse might be modifiable may not have been very difficult to arrive at, and other theories have since incorporated it, but that surely is only part of the story. I found the real impact of the story to lie in the combinatorial trick. That is, the granule cell arrangement, with associated inhibitory interneurons, had been right in front of people’s eyes ever since Cajal (modulo inhibition and excitation), but its significance had not been appreciated. Of course my theory might be wrong, but if it is right, then I would regard a major part of its contribution as being explanatory. And also, that is almost inevitable.

Many of the ideas developed in these three papers were subsequently extended and adapted to be consistent with later neurobiological discoveries. This topic was addressed in *From the Retina to the Neocortex* by L. M. Vaina (see Further Reading).

0456.007

A PIONEER OF COMPUTATIONAL NEUROSCIENCE

After the publication of these three fundamental papers, Marr moved to the Massachusetts Institute of Technology (MIT) Artificial Intelligence Laboratory where he was a visiting scientist in the group of Marvin Minsky and Seymour Papert. ‘Since the facilities and the people were really impressive’ (as he wrote to Sydney Brenner in 1973), Marr relocated from Cambridge, England to Cambridge, Massachusetts for a faculty appointment in the Department of Psychology at MIT, and in 1980 he was promoted to full professor with tenure.

0456.008

While at MIT, his decision to break with the previous research was stated clearly in a letter to Giles Brindley (written in October 1973):

0456.009

I do not expect to write any more papers in theoretical neurophysiology – at least not for a long time. I do not regard the achievements of your 1969 or my papers negligible. At the very least, they contain techniques that anyone concerned with biological computer architecture should be aware of.

This decision was motivated by his realization that, without an understanding of specific tasks and mechanisms – the issues from which his earlier theories were ‘once removed’ – any general theory would always be incomplete.

0456.010

He proposed a new methodology for understanding the brain by essentially inventing a field and a mode of study now referred to as *computational neuroscience*. In his opening remarks at a workshop organized in 1972 by Benjamin Kaminer

0456.011

at Boston University, Marr 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 favours top-down reasoning firmly supported by functional (computational) understanding, together with bottom-up work grounded in an understanding of the mechanism.

0456.012 Proposing that the primary unresolved issue in brain science was what function must be implemented and why, Marr argued fiercely against the usefulness of the theoretical approaches to brain science adopted in the early 1970s, such as the catastrophe theory pioneered by Rene Thom, and neural nets (of that time). Instead, he proposed a fundamentally novel approach to biological information processing which required that any problem must be addressed at several different levels of abstraction. What exactly was the task executed by the system? On what properties of the world could a system performing this task be expected to rely? What methods could be shown to be effective in the performance of the task? Given a particular method, what are the appropriate algorithms for implementing it? Given a particular algorithm, what neural circuitry would be sufficient to perform it? These questions formed the core of Marr's research philosophy, and they were explicitly formulated as three levels of explanation of information processing. At the highest level is a computational theory – that is, a theory of how a task could be performed. The computational theory must specify what is being computed and why it is a useful thing to compute. At the next level is a representation and an algorithm (or a set of algorithms) to achieve that representation. At the third level lies the question of how the algorithm is actually implemented in the hardware of the system. A key point in Marr's approach is that the three levels should be considered relatively independently.

0456.013 Marr's originality and depth of thinking stem from his emphasis on the computational theory level – not because it was the most important level, but because it had been generally neglected by most researchers. The computational theory of a task not only constrains the nature of the algorithm(s) for performing it, but also constrains the nature of the representation of the information at any given stage of processing. In addition, it specifies how the image is related to the outside world, by explicitly spelling out the limits on how the image can be interpreted. Knowledge of the constraints allows recovery from the image of the properties of the scene. For example, stereopsis depends on the constraint that only one point on

the retina receives light from the same source as another (unique) point on the other retina, and that the changes in disparity will be small. Although there are some possible exceptions, in general this constraint holds because the world is largely composed of smooth surfaces.

A COMPUTATIONAL THEORY OF VISION: HOW DOES THE BRAIN SEE?

This theoretical framework was the 'signature' of the research conducted in the MIT Vision Group that was formed and inspired by David Marr. The group included many talented and creative students and colleagues, such as Tomaso Poggio, Shimon Ullman, Ellen Hildreth, Eric Grimson and Keith Nishihara. Together they were seeking computational insights into the working of the visual system, and they put them to the test of implementation as computer models. Within only a few years many ground-breaking papers on computational vision had been published, including a theory of binocular stereopsis, a theory of low-level image representation, representation of direction selectivity in the cortex and a theory of the way in which shapes and actions are categorized. 0456.014

Marr's book entitled *Vision: a Computational Investigation into the Human Representation and Processing of Visual Information* (see Further Reading) is a lucid presentation of this work which proposes a general theory of the visual processing stages up to (but not including) object recognition. The framework of this theory is based on three main symbolic representations of the visual world which are created, maintained and interpreted by the process of vision. First, the *primal sketch* is mainly concerned with the description of changes in intensity of the image and their local geometry, on the grounds that intensity variations are likely to correspond to object boundaries or other physical realities. The primal sketch representation is constructed from symbolic primitives such as zero crossings, edges, contours and blobs. Secondly, the *two-and-a-half-dimensional sketch* is a viewer-centered description of the relative distances, contours and orientations of surfaces. Thirdly, the *three-dimensional model (sketch)* is an object-centered representation of objects with the goal of later allowing manipulation and recognition. This representation must be initially related to and derived from the two-and-a-half-dimensional sketch, which means that there must be a relationship between the schema of an object and the way in which 0456.015

the organization of its surfaces appears to the perceiver.

0456.016 Each of these representations is associated with algorithms used to produce them and computational theories describing specific modules in the visual system that are used to construct the sketches at each level. The idea of the vision process as a set of relatively independent modules is a powerful one from both computational and evolutionary perspectives, and some of the modules have been isolated experimentally.

DAVID MARR'S VISION

0456.017 In the winter of 1978 David Marr was diagnosed with leukemia. He died on 17 November 1980 in Cambridge, Massachusetts. His entire work provided solid proof that in behavior and brain sciences a good theory does not have to sacrifice 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.

0456.018 Through his published work, intellectual leadership, and the harmonious blend of insight, mathematical rigor and deep knowledge of neurobiology that characterizes his research, David Marr has given us a new intellectual landscape. More than two decades after his quest was cut short, research in neurobiology and cognitive sciences increasingly emphasizes the importance of elucidating the computations performed by the brain,

and the most exciting developments are those prompted (or at least accompanied) by computational theories.

Further Reading

- Marr D (1969) A theory of cerebellar cortex. *Journal of Physiology* **202**: 437–470.
- Marr D (1970) A theory for cerebral neocortex. *Proceedings of the Royal Society of London* **176**: 161–234.
- Marr D (1971) Simple memory: a theory for archicortex. *Philosophical Transactions of the Royal Society of London* **262**: 23–81.
- Marr D (1976) Early processing of visual information. *Philosophical Transactions of the Royal Society of London* **275**: 483–524.
- Marr D (1982) *Vision: a Computational Investigation into the Human Representation and Processing of Visual Information*. New York: W. H. Freeman and Company.
- Marr D and Poggio T (1976) Cooperative computation of stereo disparity. *Science* **194**: 283–287.
- Marr D and Nishihara HK (1978) Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London* **200**: 269–294.
- Marr D and Poggio T (1979) A computational theory of human stereo vision. *Proceedings of the Royal Society of London* **204**: 301–328.
- Marr D and Hildreth E (1980) Theory of edge detection. *Proceedings of the Royal Society of London* **207**: 187–217.
- Marr D and Vaina LM (1982) Representation and recognition of the movement of shapes. *Proceedings of the Royal Society of London* **214**: 501–524.
- Vaina LM (1991) *From the Retina to the Neocortex. Selected Papers of David Marr*. Boston, MA: Birhauser, Springer Verlag.

Keywords:

biography; theoretical neurobiology; computational vision

Article 456 David C. Marr

1. Please supply copy for a Definition.
2. Because this is a Level 1 article (with a Further Reading section but no References section), all of the references cited in the text have been removed, and the heading “References” has been changed to “Further Reading”.
3. Introduction, line 9. Is it possible to give date when Marr received his PhD?
4. Contributions to theoretical neuroscience, para 2, line 8. Is “at the level of the parallel fiber–Purkinje cell synapse” OK now?
5. Same section, para 5, quoted extract. Is “the idea that the parallel fiber–Purkinje cell synapse might be modifiable” OK now?
6. A pioneer of computational neuroscience, para 2, line 4. Please rephrase “I do not regard the achievements of your 1969 or my papers negligible”.
7. Same section, last para, line 9. Is “the image of the properties of the scene” OK now?
8. A computational theory of vision, para 2, line 12. Is “the two-and-a-half-dimensional sketch” correct?
9. Same section, para 2, third line from end. Is “which means that there must be a relationship” OK now?