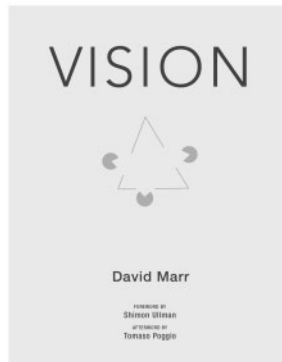
doi:10.1093/brain/awr013

BRAIN
A JOURNAL OF NEUROLOGY

BOOK REVIEW

David Marr's Vision: *floreat* computational neuroscience

VISION: A COMPUTATIONAL INVESTIGATION INTO THE HUMAN REPRESENTATION AND PROCESSING OF VISUAL INFORMATION
By *David Marr* 2010. London and Cambridge, MA: MIT Press
Price: £22.95/$30
ISBN: 978-0-262-51462-0



This is a reprint of David Marr's 1982 book. A foreword placing the book in its historical context is added by Shimon Ullman, and an afterword by Tomaso Poggio is added on some of the themes in the book. David Marr was one of the originators of computational neuroscience, and the useful re-publication of this book enables us to assess how this field is developing, and to put David Marr's contributions into perspective. David Marr (1945–80) obtained a First Class degree in Mathematics at the University of Cambridge in 1966; and was sufficiently interested in how the brain works to attend the Part II undergraduate courses in physiology and psychology of the Natural Sciences Tripos. (David was not experienced in practical classes, and happened to be paired with Barbara Rolls, the first female PhD student in physiology at Cambridge, who also sat in on the practical classes and provided expertise partly as a result of her training with Alan Epstein at the University of Pennsylvania.) One of the lecturers in physiology was Giles Brindley, who was interested in vision (as were many of the other members of the Department, including Horace Barlow, Fergus Campbell, William Rushton and John Robson) and in synaptic physiology. [Giles Brindley's Physiology of the Retina and Visual Pathway (Physiological Society Monograph No. 6, Edwin Arnold, London) appeared in 1970.] Giles Brindley published a paper on how different classes of synapses might show plasticity and contribute to learning in neural networks (Brindley, 1969). These lectures and this work stimulated David's thinking about synaptic modification and its role in systems in the brain that learn. This led to three seminal papers: on the hippocampus (Marr, 1971), the cerebellum (Marr, 1969) and the neocortex (Marr, 1970). David's theory of the hippocampus was influenced too at the systems level by Larry Weiskrantz's Part II lectures in psychology, which treated topics such as memory and emotion (Weiskrantz, 1956, 1968*a*, *b*; Weiskrantz and Saunders, 1984). The same Part II lectures also influenced my own research on memory, emotion and vision (Rolls, 2005, 2008).

One important property of David Marr's approach at this time was the move to take into account the quantitative network architecture of the brain system being modelled—the hippocampus, cerebellum and neocortex (Marr, 1969, 1970, 1971)—to produce a quantitative theory. This has proven to be very important in subsequent computational neuroscience approaches to memory, vision, attention and decision making (Rolls and Treves, 1998; Rolls and Deco, 2002; Rolls, 2008; Rolls and Deco, 2010). However, neuroscience was insufficiently advanced in the 1970s for David Marr to put his theories to empirical test. Nonetheless, he did try—working for example with John Eccles (Eccles *et al*., 1967) to test the prediction that the cerebellar parallel fibre to Purkinje cell synapses would modify associatively with the inferior olive/climbing fibre input to the Purkinje cell. They were not able to confirm this prediction, perhaps in part because the climbing fibre input was stimulated at much higher rates than these fibres are now known to fire naturally (in the range of 0–10 spikes/s). But this fundamental tenet and prediction of the theory of learning that occurs at these synapses was subsequently confirmed (Ito, 1984). Partly because of this difficulty in testing his neural network theories of cortical structure in the 1970s, David Marr chose to move to a different level of investigation in which computations being performed were studied, and tested by psychophysics, rather than being modelled at the level of their implementation in the brain. It is at this level that his 1982 book, *Vision*, is written. David performed the research for the book at the Massachusetts Institute of Technology (MIT) where he had moved in 1973, partly, I was told, because MIT could provide a teletype in his bedroom with a connection to a large computer on the campus. (The University of Cambridge was, however, quite advanced in computing at the time, and I remember while an undergraduate helping to dismantle EDSAC, one of the first large British computers; it was described as rather unreliable, with thousands of triode valves to implement flip-flops.)

Vision thus describes a computational approach to human vision. The first part of the book is concerned with early visual processing (called the primal sketch by Marr), including edge detection, stereopsis, directional selectivity, shape contours, surface texture and shading. These are areas in which Marr made important contributions. Chapter 4 (From Images to Surfaces) describes Marr's 2½-D sketch which is a viewer-centred representation of the visible surfaces based on the results of early visual processing. An example of what he meant by a 2½-D sketch is illustrated by his Figure 3-12 (Fig. 1). This is an important advance, for it goes beyond the concept of segmentation of the visual scene into objects, as an important step in the early analysis of vision, to focus instead on using all the information that is available to represent the surfaces that are actually visible and their depths from the observer as a precursor to analysis at a later stage. This again is useful, for computer vision approaches have great difficulty in segmenting whole scenes into objects using simple early vision algorithms. David Marr used the subjective contours visible in his Figure 2-6 (Fig. 2) to emphasize the importance of representing contour and depth even when there is no direct visual evidence for them.

Chapter 5 is concerned with Representing Shapes for Recognition. Marr's 3D sketch is described here, and involves representing parts of objects and their syntactic relation to each other (e.g. the fingers are attached to the palm and not to the elbow or trunk). Object recognition was approached by attempting to match a syntactic description of an object with the stored syntactic descriptions of all objects. The approach has the aim of producing view-invariant object representation by specifying the parts of an object that are visible, and their relation to each other, which provides a view invariant description of an object suitable for view invariant object recognition. Marr's famous example was the representation of the human body as a set of interlinked generalized cones (Marr and Nishihara, 1978), with the approach illustrated in his Figure 5-3 (Fig. 3). As a theory of object recognition in the brain, this has proved intractable. It is very hard to extract all the cylinders or shape components that describe objects from a complex scene; very hard to know which shape primitives belong to a single object; very hard to represent the syntactic relations in a neuronal network; and very hard to perform the look-up of a syntactic description of a visible object with all possible stored 3D sketches of objects to perform object recognition (Rolls, 2008). Marr, in fact, recognized that his approach would have been strengthened and would perhaps have changed with time, writing poignantly in the summer of 1979 in the preface to Vision: 'events happened which forced me to write this book a few years earlier than I had planned'. (David Marr died of leukaemia in 1980 at the age of 35; and *Vision* was published posthumously in 1982.)

Instead, theory and closely linked empirical research suggest that the brain takes a very different approach to invariant object recognition—recognition that is invariant with respect to the position of the object on the retina, its size and even its view (Rolls, 1992, 2000, 2008, 2011). The present understanding is that the brain uses associative learning that involves temporal and spatial continuity (which is a property of objects as they transform in the

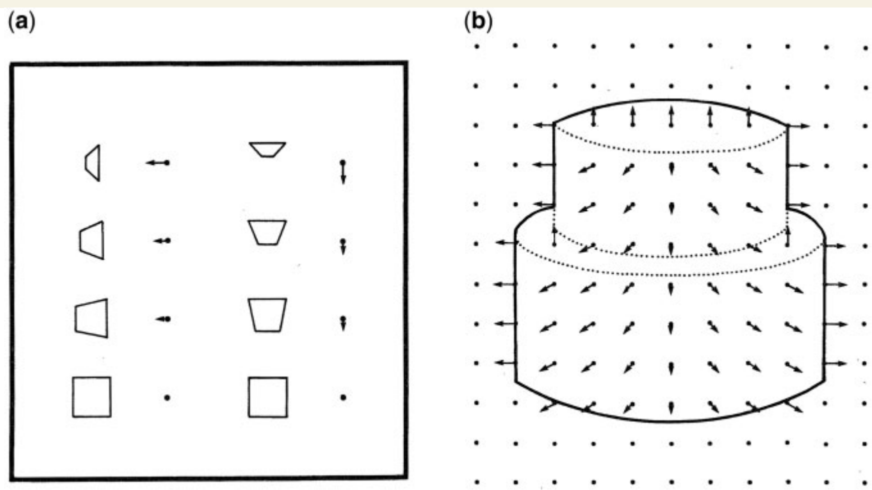


Figure 1 Figure 3-12: illustration of the 2½-D sketch. (**a**) The perspective views of small squares placed at various orientations to the viewer are shown. The dots with arrows symbolically represent the orientations of such surfaces. (**b**) This symbolic representation is used to show the surface orientations of two cylindrical surfaces in from of a background orthogonal to the viewer. The full 2½-D sketch would include rough distances to the surfaces as well as their orientations; contours where surface orientations change sharply, which are shown dotted; and contours where depth is discontinuous (subjective contours), which are shown with full lines. (Reprinted by permission from Marr and Nishihara. Representation and recognition of the spatial organization of three-dimensional shapes. Proc R Soc Lond B 1978; 200: 269–294.)

world) at several stages of the cortical hierarchy, from the primary visual cortex (V1) to the inferior temporal visual cortex, to build what are effectively view-based representations of parts and of whole objects that are then associated together to form representations, which can be accessed associatively in a view-invariant way. That approach has also been accepted by other investigators, including Tomaso Poggio (Riesenhuber and Poggio, 1999, 2000; Serre *et al.*, 2007), who worked with David Marr. The process is simplified and made tractable by processing only small parts of a scene at any time—that which is close to the fovea and fixated at any one time. The receptive fields even become smaller, less than 10° in diameter, in complex natural scenes due to lateral inhibition; and this is part of the solution to the binding problem which is thereby greatly reduced as only one or a few objects close to the fovea are processed at any one time by the inferior temporal visual cortex where object recognition is represented (Rolls, 2008).

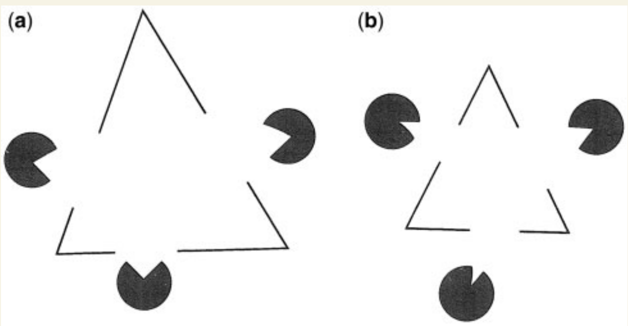


Figure 2 Figure 2-6: subjective contours. The visual system apparently regards changes in depth as so important that they must be made explicit everywhere, including places where there is no direct visual evidence for them.

One of the key issues addressed by Marr in *Vision* is the level of analysis that is used in computational neuroscience. Marr favoured the top level, the *computational theory* level—what is the goal of the computation; why is it appropriate; and what is the logic of the strategy by which it can be carried out? He distinguished this from the second level, the *representation and algorithm*—how can this computational theory be implemented? In particular, what is the representation for the input and output, and what is the algorithm for the transformation? His third level is *hardware implementation*—how can the representation and algorithm be realized physically? In his earlier work on the cerebellar, neocortical and hippocampal theories (Marr, 1969, 1970, 1971), he had included much on the third level, implementation in the brain, and this was being used to help constrain the computational theory. But, perhaps partly for the reasons given above, in *Vision* he strongly favoured the computational theory approach, suggesting that one should start here.

However, when understanding the cortical mechanisms of vision, what is found neurophysiologically (Hubel and Wiesel, 1968; Rolls, 2000, 2008, 2011) and in terms of the neuronal network architecture in the brain provides very important constraints on the theory, whether this is of vision, memory, attention or decision making (Rolls and Treves, 1998; Rolls and Deco, 2002, 2010; Rolls, 2008). Thus a more modern approach, which is making very fast progress at present, is to combine empirical neurophysiological and neuroanatomical data with approaches

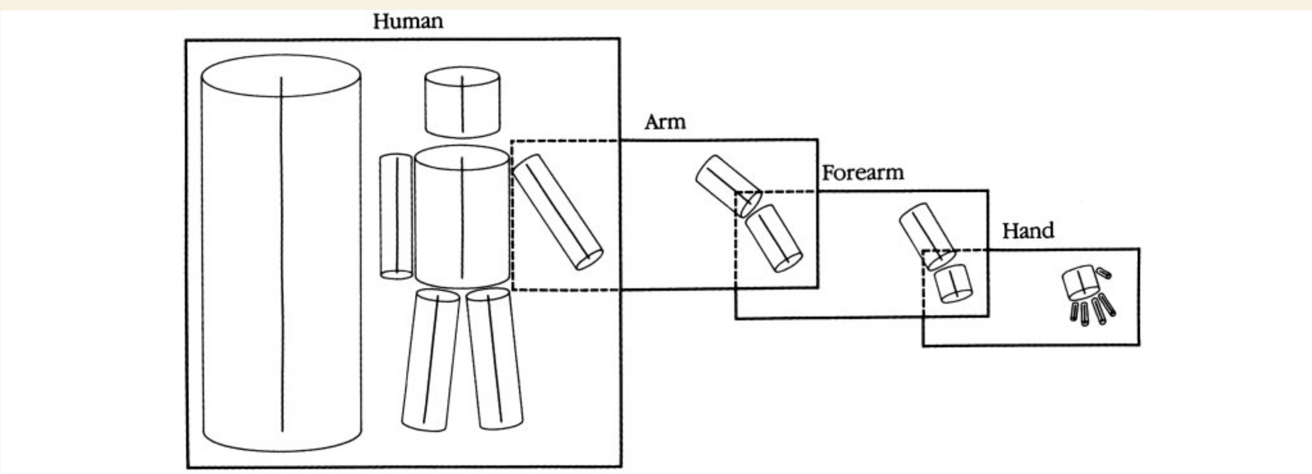


Figure 3 Figure 5-3: this diagram illustrates the organization of shape information in a 3D model description of an object based on generalized cone parts. Each box corresponds to a 3D model, with its model axis on the left side of the box and the arrangement of its component axes on the right. In addition, some component axes have 3D models associated with them, as indicated by the way the boxes overlap. The relative arrangement of each model's component axes, however, is shown improperly, since it should be in an object-centred system rather than the viewer-centred projection used here. The important characteristics of this type of organization are: (i) each 3D model is a self-contained unit of shape information and has a limited complexity; (ii) information appears in shape contexts appropriate for recognition (the disposition of a finger is most stable when specified relative to the hand that contains it); and (iii) the representation can be manipulated flexibly. This approach limits the representation scope, however, since it is only useful for shapes that have well-defined 3D model decompositions. (Reprinted by permission from Marr and Nishihara. Representation and recognition of the spatial organization of three-dimensional shapes. Proc R Soc Lond B 1978; 200: 269–294).

that produce and test theories of how the brain computes (Rolls, 2008). In turn, this strategy is informing a new approach to neurological psychiatry that seeks to understand certain disorders of brain function (including schizophrenia and obsessive compulsive disorder) by analysing the stochastic dynamics and stability of cortical systems (Rolls, 2008; Rolls *et al.*, 2008*a*, *b*; Rolls and Deco, 2010); and this again relies on combining theory with empirical research. Marr was certainly right in the following: without theoretical approaches being part of how we understand brain function, we will never understand how vision works, or for that matter memory, attention, decision making and some neuropsychiatric disorders of cortical function.

Shimon Ullman in his foreword comments that research monographs age quickly, but that because Marr treated broad problems such as how the brain and its functions can be studied, one may still enjoy the book and appreciate his creativity, intellectual power and ability to integrate insights and data from the fields of neuroscience, psychology and computation. Ullman notes that the central role of invariant 3D models such as that proposed by Marr has been challenged by subsequent psychophysical and computational studies, which have moved towards an alternative approach to recognition, based on describing the possible image appearances of an object rather than its invariant 3D structure, consistent with the type of model described above (Rolls, 1992, 2008; Serre *et al.*, 2007).

Tomaso Poggio, in his afterword, notes that Marr's *Vision* played a key role in the beginning and rapid growth of the field of computational neuroscience. Poggio does agree though that it is now time to re-emphasize the connections between the levels of analysis described by Marr, if we want to make progress in computational neuroscience; and he has an interesting account of how the original 'manifesto' for their computational approach to brain function was developed. Poggio indeed makes a salutary point about theory and explanation in connection with oscillations in the brain, where Marr's message may sometimes be lost. Poggio notes that 'an explanation of the biophysics of oscillations in the neural activity of cortical areas appears to be regarded in several papers as a full explanation in itself, whereas, in the spirit of computational neuroscience, one must also eventually understand what is the computational role of oscillations and what is the algorithm that controls them. In other words, oscillations may be a symptom or the mechanism of attention, but which computation is actually performed by oscillations?' That challenge is now being addressed (Deco and Rolls, 2011).

On re-viewing *Vision*, one is struck by the deep almost philosophical but in fact computational considerations that Marr brought to understanding brain function. He not only does this, but also has an interesting Chapter 7 (A Conversation) where he discusses with himself in almost Platonic dialogue, putting objections to, and justifications for, the computational approach he takes. His reflective and penetrating thought is one of his lasting contributions, as is his approach to computational neuroscience, in which he was one of the pioneers: *floreat* computational neuroscience.

Edmund T. Rolls
Oxford Centre for Computational Neuroscience, Oxford, UK.
www.oxcns.org; Edmund.Rolls@oxcns.org

References

Brindley GS. Nerve net models of plausible size that perform many simple learning tasks. Proc R Soc Lond B Biol Sci 1969; 174: 173–91.
Deco G, Rolls ET. How oscillations add to firing rates. Trends Neurosci 2011.
Eccles JC, Ito M, Szentágothai J. The cerebellum as a neuronal machine. New York, Heidelberg: Springer; 1967.
Hubel DH, Wiesel TN. Receptive fields and functional architecture of macaque monkey striate cortex. J Physiol 1968; 195: 215–43.
Ito M. The cerebellum and neural control. New York: Raven Press; 1984.
Marr D. A theory of cerebellar cortex. J Physiol 1969; 202: 437–70.
Marr D. A theory for cerebral neocortex. Proc R Soc Lond B Biol Sci 1970; 176: 161–234.
Marr D. Simple memory: a theory for archicortex. Phil Trans Roy Soc Lond B 1971; 262: 23–81.
Marr D, Nishihara HK. Representation and recognition of the spatial organisation of three dimensional shapes. Proc Royal Soc Lond B 1978; 200: 269–94.
Riesenhuber M, Poggio T. Hierarchical models of object recognition in cortex. Nat Neurosci 1999; 2: 1019–25.
Riesenhuber M, Poggio T. Models of object recognition. Nat Neurosci 2000; 3 (Suppl): 1199–204.
Rolls ET. Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas. Phil Trans Roy Soc Lond B 1992; 335: 11–21.
Rolls ET, Treves A. Neural networks and brain function. Oxford: Oxford University Press; 1998.
Rolls ET. Functions of the primate temporal lobe cortical visual areas in invariant visual object and face recognition. Neuron 2000; 27: 205–18.
Rolls ET, Deco G. Computational neuroscience of vision. Oxford: Oxford University Press; 2002.
Rolls ET. Emotion explained. Oxford: Oxford University Press; 2005.
Rolls ET. Memory, attention, and decision-making: a unifying computational neuroscience approach. Oxford: Oxford University Press; 2008.
Rolls ET, Loh M, Deco G. An attractor hypothesis of obsessive-compulsive disorder. Eur J Neurosci 2008a; 28: 782–93.
Rolls ET, Loh M, Deco G, Winterer G. Computational models of schizophrenia and dopamine modulation in the prefrontal cortex. Nat Rev Neurosci 2008b; 9: 696–709.
Rolls ET, Deco G. The noisy brain: stochastic dynamics as a principle of brain function. Oxford: Oxford University Press; 2010.
Rolls ET. Face neurons. In: Calder AJ, Rhodes G, Johnson MH, Haxby JV, editors. The handbook of face perception, Ch 4. Oxford: Oxford University Press; 2011, p. 51–75..
Serre T, Wolf L, Bileschi S, Riesenhuber M, Poggio T. Robust object recognition with cortex-like mechanisms. IEEE Trans Pattern Anal Mach Intell 2007; 29: 411–26.
Weiskrantz L. Behavioral changes associated with ablation of the amygdaloid complex in monkeys. J Comp Physiol Psychol 1956; 49: 381–91.
Weiskrantz L. Emotion. In: Weiskrantz L, editor. Analysis of behavioural change. New York and London: Harper and Row; 1968a. p. 50–90.
Weiskrantz L. Experiments on the r.n.s. (real nervous system) and monkey memory. Proc R Soc Lond B Biol Sci 1968b; 171: 335–52.
Weiskrantz L, Saunders RC. Impairments of visual object transforms in monkeys. Brain 1984; 107: 1033–72.