

CROSS TALK

Marr's three levels of analysis are useful as a framework for neuroscience

Máté Lengyel^{1,2} ¹Computational and Biological Learning Lab, Department of Engineering, University of Cambridge, Cambridge, UK²Center for Cognitive Computation, Department of Cognitive Science, Central European University, Budapest, Hungary

Email: m.lengyel@eng.cam.ac.uk

Handling Editors: Francisco Sepúlveda & Vaughan Macefield

In September 2019, I had the pleasure of attending a symposium (Galliano et al., 2019) commemorating the legacy of David Marr, one of the pioneers of computational neuroscience – the field concerned with developing mathematical models and analysis tools for studying the nervous system. At the symposium, I also participated in a live debate about the merits of one particular contribution that Marr made to the field: the so-called three levels of analysis (Marr, 1982). This CrossTalk article is an extended summary of the arguments I made there for why I find this contribution still valuable, more than 40 years after its publication.

It is probably fair to say that all students of computational neuroscience come across David Marr's legacy at some point during their education. In fact, this encounter usually happens right towards the very beginning of most of our introductory courses. This is because, besides his numerous seminal *scientific* contributions to the field (e.g. Marr, 1969, 1970, 1971; Marr & Hildreth, 1980), the one contribution of Marr that has become a staple of our curricula is actually a *metascientific* one (or as Marr called it: 'philosophical preliminaries'). It is concerned with 'the three levels at which any machine carrying out an information-processing task must be understood' (Marr, 1982) – where Marr generously included the nervous system in his definition of a 'machine'.

Marr's three levels

Quoting from Marr (1982), the three levels of analysis answer the following distinct questions:

Computational theory: 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?

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?

Hardware implementation: How can the representation and algorithm be realized physically?

Let me illustrate these (perhaps somewhat abstract-sounding; c.f. 'philosophical preliminaries') levels via two concrete, Nobel Prize-winning examples of neurophysiological phenomena, which will be intimately familiar to most readers of this journal (not the least because the original discoveries were reported here). While both these original discoveries predate Marr's conceiving of his levels, a large body of important work followed later about the same phenomena – and, as I argue below, Marr's levels help us also appreciate the importance of these later contributions.

- **The firing of action potentials.** Hodgkin's and Huxley's seminal work revealed the dynamical mechanisms in the membrane (i.e. the opening and closing of ion channels) that underlie one of the most important signals in the nervous system: the action potential (Hodgkin et al., 1952; Hodgkin & Huxley, 1952a, 1952b, 1952c). That work laid the foundations of modern electrophysiology and (together with Rall's cable theory; Rall, 1959) of our understanding of how the concerted action of ion channels leads to the generation and propagation of electrical signals in a neuron. All this work, thus, sits firmly at Marr's *implementational level*. However, abstracting away from the detailed biophysical implementation, one can

also ask what is the overall mapping between the inputs (synaptic currents) and output of a neuron (somatic membrane potential, including the firing of action potentials)? One could even argue that as far as the operation of a neural circuit is concerned it is *only* this phenomenological input–output mapping that matters (Zador, 2000). There has, thus, been a plethora of influential studies at the algorithmic level aimed at characterizing this mapping (Beniaguev et al., 2021; Poirazi et al., 2003; Ujfalussy et al., 2018). Finally, at the computational level, one can also ask questions about the functional (dis)advantages of the action potential generating mechanisms revealed by Hodgkin and Huxley: why do neurons use those particular ion channels with those particular kinetics? Interestingly, comparatively less work studied these computational-level questions. Nevertheless, studies on the energy consumption, and indeed, efficiency of the nervous system suggest that in many cases, especially in the mammalian cortex (though perhaps not in the squid giant axon, where Hodgkin and Huxley first described them) these mechanisms might be optimized to consume minimal energy while still reliably maintaining the shape of the action potential (Alle et al., 2009; Sterling & Laughlin, 2015).

- **The oriented receptive fields of simple cells in the primary visual cortex.** The original description of receptive fields by Hubel and Wiesel (1959) is a *bone fide* example of a contribution at the *representational level*. Receptive fields describe (with some approximation) how a cell responds as a function of the visual image displayed to the animal. Thus, they define a transformation from input (visual image) to output (neural responses). Of course, what receptive fields do not reveal as such is how this transformation comes about in the circuitry of the visual system. This is most apparent in the fact that

they identify the input as the image displayed to the animal, while cells in the primary visual cortex (V1) do not receive such input directly. Indeed, starting with early work by Hubel and Wiesel (1962) themselves, a great deal of effort has gone into understanding how these receptive fields are realized by the circuit – thus answering questions at the level of *hardware implementation*. Whatever the detailed circuit mechanisms underlying them, oriented receptive fields seem to be intuitively useful as they detect edges – of which there are many in our visual environment (Geisler et al., 2001). To formalize this link between environment and neural behaviour more rigorously, rather than just relying on intuition, one needs a *computational theory*. Such influential theories have shown that these receptive fields serve important adaptive functions, in terms of reducing the redundancy (Barlow, 1961; Bell & Sejnowski, 1997), or increasing the sparseness (Olshausen & Field, 1996), of signals along the visual pathway, and – more recently – producing feature detectors that greatly aid object recognition (Yamins et al., 2014).

Usefulness

The kind of ‘philosophical preliminaries’ epitomized by Marr’s three levels is often eschewed in our science classes. Nevertheless, taking one step back and forcing ourselves to think more broadly about the overall approach we are taking to study a problem (such as ‘how does the brain work?’), and considering alternative angles, can only do good in the long run. Indeed, there is a venerable history of such epistemological insights enriching and contributing to the progress of even some of the ‘hardest’ sciences, including modern physics (Norton, 2010).

More pragmatically, it can be liberating and inspiring at the same time to realize that in order to study any particular phenomenon, one does not need to answer *all* these types of questions at once. Indeed, contributing to one level at a time can already make a significant contribution. This insight becomes especially (and somewhat painfully) relevant when the need arises to justify the particular

approach that one chooses to follow in studying a phenomenon – and this approach does not happen to coincide with the majority’s choice. Historically, many key phenomena in neuroscience, including the examples above, were first studied – with great success – at the implementational or representational level. However, these early successes can sometimes create the false association as if studying these phenomena was equivalent to asking the particular questions relevant at any one of these levels. Thus, computational-level analyses often bore the burden of having to justify their existence as they entered an arena already populated by influential implementational or representational level analyses.

The two examples discussed above are good points in case. A computational level theory of V1 simple cells may (and, emboldened by Marr, need) say nothing about the mechanisms underlying those receptive fields: the synaptic cascades of the early visual system upstream of V1, the circuit mechanisms within V1 or feedback from areas downstream from V1. Such a theory may, thus, not offer predictions for the consequences of a particular pharmacological or optogenetic manipulation. Instead, by examining the potential functions of these receptive fields, they may make predictions about how those V1 cell responses depend on the statistics of natural images (Olshausen & Field, 1996), and how V1 is being adapted to those statistics (Berkes et al., 2011). Similarly, an algorithmic or computational level analysis of dendritic integration may remain agnostic as to the individual ion channels and currents contributing to a cell’s response, but may help us better understand the functions of those integrative properties in the context of the computations the whole circuit performs (Kim et al., 2023; Poirazi et al., 2003; Ujfalussy et al., 2018). Given the continued success of neurophysiological investigations typically focusing on *mechanisms*, it may at first seem heretic for a theory not to make implementational level predictions. Yet, the examples above make it clear that the levels above the implementation can also offer valuable insights. Marr’s distinction between different levels of analysis greatly helps us articulate these insights, or even realize that relevant questions at these higher levels exist, and appreciate what addressing these questions brings to the table of neuroscience at which we all sit.

Of course, as history goes, tides shift and there are areas in neuroscience in which computational and algorithmic approaches have taken precedence over implementational ones. For example, the study of phasic dopaminergic signals in the midbrain has long been dominated by the reward prediction error framework (Schultz et al., 1997), which is at heart a computational and algorithmic theory (and, incidentally, a landmark result of computational and systems neuroscience). Only recently detailed implementational-level questions have started being asked and answered about it (with the advent of modern experimental approaches, including optogenetics; Eshel et al. 2015). Nevertheless, even these implementational-level studies often build on the foundations of the original algorithmic approach – representing a beautiful example of the ultimate synergy of Marr’s levels necessary for a (more) complete understanding of the nervous system.

Pluralism versus primacy

Marr himself strongly argued for an approach to understanding vision (note the title of the book, ‘Vision’, in which the levels of analysis were introduced, Marr 1982) that was deeply rooted in computational-level analyses. This can easily be mistaken for advocating for the primacy of computational-level analyses. That this was not Marr’s intention is clear from the way he prepares ground for his three levels:

Such an analysis [at the computational level – M.L.] does not usurp an understanding at the other levels – of neurons or computer programs [representation and algorithm – M.L.] – but it is a necessary complement to them, since without it there can be no real understanding of the function to all those neurons.

As well as from the ‘Synopsis’ of his book:

I have laid particular stress on the level of computational theory, not because I regard it as inherently more important than the other two levels – the real power of the approach lies in the integration

Table 1. Levels of analysis according to various cognitive and neuroscientists

Marr	Chomsky	Pylyshyn	Rumelhart and McClelland	Newell	Anderson	Dayan and Abbott
Competence						
Computational theory		Semantic level		Knowledge level	Rational level	Interpretive
Representation and algorithm	Performance	Algorithm	Macrotheory /rules	Programme symbol level	Algorithm	Descriptive
		Functional architecture	Microtheory PDP models	Register transfer level	Implementation	
Hardware implementation		Biological level		Device	Biological	Mechanistic

Reproduced from Anderson (1990) and extended with the last column. References for these classifications from left to right are: Marr (1982), Chomsky (2014), Pylyshyn (1984), Rumelhart and McClelland (1985), Newell (1982), Anderson (1990) and Dayan and Abbott (2005). PDP stands for parallel distributed processing. The analogies between levels that appear in register across different classifications are sometimes loose, especially between those of Dayan and Abbott (2005) and the rest.

of all three levels of attack – but because it is a level of explanation that has not previously been recognized and acted upon. It is, therefore, probably one of the most difficult ideas for newcomers to the field to grasp, and for this reason alone its importance should not be understated in any introductory book, such as this is intended to be.

Thus, it is clear that Marr was actually strongly pro-pluralism of approaches, and his emphasis on the computational level was simply due to the historical context of vision research at the time which (Marr felt) was unduly skewed towards algorithmic/representational and implementational-level analyses – an imbalance that Marr thought needed to be rectified. Once again, his levels of analyses were useful in identifying that there was an imbalance in approaches in the first place, and then also for articulating the kind of analyses that were necessary to fill the gap that the then-dominant approaches left.

Uniqueness

There is a potential problem with any classification (at least in 'soft subjects', such as neuroscience): there are bound to be

borderline cases that do not fit neatly into any one of the categories. Marr's levels are no exception: there may be no unique assignment of a given study or approach to any one of his levels. However, this need not be a problem: in fact, some of the most exciting developments in computational neuroscience often happen by bridging different levels – and acknowledging this bridge-building by noting that they belong to more than just one of Marr's levels can help us better appreciate their accomplishments (Doiron & Lengyel, 2019).

More broadly, it is perhaps relatively little contested at this point that distinguishing between different levels of analysis is useful for structuring the field of neuroscience and guiding our thoughts and discussions about the virtues and pitfalls of existing as well as future studies. What may be less obvious is whether the specific set of three levels that Marr proposed are uniquely the best for this purpose. Indeed, there have been several similar classifications proposed over the history of neuroscience and cognitive science (Table 1; Anderson, 1990). There are two obvious conclusions one can draw from the long list of similar classifications. First, there is a need for such classifications because distinguishing between different levels of description can

be useful in several ways – some of which I discussed above. Second, what is also immediately apparent is that there is no universal agreement about the 'right' set of levels, or even the right way of defining these levels. I will not attempt a detailed metaanalysis of the levels of analyses in neuroscience here, and instead just note that despite some obvious differences, it is clear that these different classifications have a lot in common and are trying to achieve very similar goals. Marr's classification may not be the best – for example, I personally favour Dayan and Abbott's distinction of interpretive, descriptive and mechanistic approaches (Table 1, last column). Nevertheless, Marr's levels are *good enough* and, until a better classification of the same sort becomes similarly widespread in use, they are certainly better than having no such classification and thus no shared common language to express the fundamentally different, but equally valid and complementary contributions offered by the different levels of analysis of the nervous system.

In summary, Marr's levels are useful, even if imperfect and not unique, and do not imply the primacy of any one of the levels – on the contrary, they remind us of the importance of studying the nervous system at each of these levels.

Call for comments

Readers are invited to give their views on this and the accompanying Cross Talk articles in this issue by submitting a brief (250 words) comment. Comments may be submitted up to 6 weeks after publication of the article, at which point the discussion will close and the Cross Talk authors will be invited to submit a 'LastWord'. Please email your comment, including a title and a declaration of interest, to jphysiol@physoc.org. Comments will be moderated, and accepted comments will be published online only as 'supporting information' to the original debate articles once discussion has closed.

References

- Alle, H., Roth, A., & Geiger, J. R. (2009). Energy-efficient action potentials in hippocampal mossy fibers. *Science*, **325**(5946), 1405–1408.
- Anderson, J. R. (1990). *The adaptive character of thought*. Psychology Press.
- Barlow, H. B. (1961). Possible principles underlying the transformations of sensory messages. In: W. Rosenblith (Ed.), *Sensory communication* (pp. 217–234). MIT Press.
- Bell, A. J., & Sejnowski, T. J. (1997). The "independent components" of natural scenes are edge filters. *Vision Research*, **37**(23), 3327–3338.
- Beniagiev, D., Segev, I., & London, M. (2021). Single cortical neurons as deep artificial neural networks. *Neuron*, **109**(17), 2727–2739.e3.
- Berkes, P., Orbán, G., Lengyel, M., & Fiser, J. (2011). Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science*, **331**(6013), 83–87.
- Chomsky, N. (2014). *Aspects of the theory of syntax*. MIT Press.
- Dayan, P., & Abbott, L. F. (2005). *Theoretical neuroscience*. MIT Press.
- Doiron, B., & Lengyel, M. (2019). Editorial overview: Computational neuroscience. *Current Opinion in Neurobiology*, **58**, iii–vii.
- Eshel, N., Bukwich, M., Rao, V., Hemmelder, V., Tian, J., & Uchida, N. (2015). Arithmetic and local circuitry underlying dopamine prediction errors. *Nature*, **525**(7568), 243–246.
- Galliano, E., Glynn, D., Edgley, S., Paulsen, O., Lengyel, M., Eglén, S., Otis, T., & Tripodi, M. (2019). *David Marr—50 years on*. <https://www.neuroscience.cam.ac.uk/events/ABC2019/pages/about-marr.html>.
- Geisler, W. S., Perry, J. S., Super, B. J., & Gallogly, D. P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research*, **41**(6), 711–724.
- Hodgkin, A. L., & Huxley, A. F. (1952a). Currents carried by sodium and potassium ions through the membrane of the giant axon of *Loligo*. *The Journal of Physiology*, **116**(4), 449–472.
- Hodgkin, A. L., & Huxley, A. F. (1952b). The components of membrane conductance in the giant axon of *Loligo*. *The Journal of Physiology*, **116**(4), 473–496.
- Hodgkin, A. L., & Huxley, A. F. (1952c). The dual effect of membrane potential on sodium conductance in the giant axon of *Loligo*. *The Journal of Physiology*, **116**(4), 497–506.
- Hodgkin, A. L., Huxley, A. F., & Katz, B. (1952). Measurement of current-voltage relations in the membrane of the giant axon of *Loligo*. *The Journal of Physiology*, **116**(4), 424–448.
- Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology*, **148**(3), 574–591.
- Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, **160**(1), 106–154.
- Kim, Y. J., Ujfalussy, B. B., & Lengyel, M. (2023). Parallel functional architectures within a single dendritic tree. *Cell Reports*, **42**(4), 112386.
- Marr, D. (1969). A theory of cerebellar cortex. *The Journal of Physiology*, **202**(2), 437–470.
- Marr, D. (1970). A theory for cerebral neocortex. *Proceedings of the Royal Society B: Biological Sciences*, **176**(1043), 161–234.
- Marr, D. (1971). Simple memory: A theory for archicortex. *Philosophical Transactions of the Royal Society B*, **262**(841), 23–81.
- Marr, D. (1982). *Vision: A computational approach*. Freeman & Co.
- Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society B: Biological Sciences*, **207**(1167), 187–217.
- Newell, A. (1982). The knowledge level. *Artificial Intelligence*, **18**, 87–127.
- Norton, J. D. (2010). How Hume and Mach helped Einstein find special relativity. In M. Domski & M. Dickson (Eds.), *Discourse on a new method: Reinvigorating the marriage of history and philosophy of science* (pp. 359–387). Open Court Publishing.
- Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, **381**(6583), 607–609.
- Poirazi, P., Brannon, T., & Mel, B. W. (2003). Pyramidal neuron as two-layer neural network. *Neuron*, **37**(6), 989–999.
- Pylyshyn, Z. W. (1984). *Computation and cognition*. MIT Press.
- Rall, W. (1959). Branching dendritic trees and motoneuron membrane resistivity. *Experimental Neurology*, **1**, 491–527.
- Rumelhart, D. E., & McClelland, J. L. (1985). Levels indeed! a response to Broadbent. *Journal of Experimental Psychology: General*, **114**, 193–197.
- Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, **275**(5306), 1593–1599.
- Sterling, P., & Laughlin, S. (2015). *Principles of neural design*. MIT Press.
- Ujfalussy, B. B., Makara, J. K., Lengyel, M., & Branco, T. (2018). Global and multiplexed dendritic computations under *in vivo*-like conditions. *Neuron*, **100**(3), 579–592.
- Yamins, D. L., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, **111**(23), 8619–8624.
- Zador, A. M. (2000). The basic unit of computation. *Nature Neuroscience*, **3**, 1167–1167.

Additional information

Competing interests

The author declares no competing interests.

Author contributions

Sole author.

Funding

None.

Acknowledgements

I thank the organizers and participants of the Marr Symposium for the original discussion, my partner in debate, Jonathan Pillow for his inspiring arguments and Zs. Novák for useful references to the influence of epistemology on physics. This work was supported by the Wellcome Trust (Investigator Award in Science 212262/Z/18/Z), the Human Frontiers Science Programme (Research Grant RGP0044/2018) and the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (Grant Agreement No. 726090).

Keywords

action potential, behaviour, computational neuroscience, dendrite, neural circuits, neural coding, receptive field, vision

Supporting information

Additional supporting information can be found online in the Supporting Information section at the end of the HTML view of the article. Supporting information files available:

Peer Review History