Chapter on Vision for *Routledge Handbook of the Computational Mind*

M. Chirimuuta
History & Philosophy of Science, University of Pittsburgh

1. Preamble: Is the Eye like a Camera?

In today's textbooks, and for centuries before now, it has been commonplace to explain the workings of the mammalian eye as being similar to those of a non-biological camera. The same principles of optics are exemplified in each kind of device, and it has tempted some to say that the eye *is* a fleshy, gelatinous, photographic machine. So from our vantage point it is surprising that before Kepler's discovery of the optical laws of image formation in the eye (living or dead) and *camera obscura*, this comparison was not readily available to early theorists of vision. Ibn al-Haytham ("Alhazen") was the first to apply inferentialist notions to the theory of vision, prefiguring Helmholtz and also current computational accounts. However, he did not explain the workings of the eye in terms of optical laws that hold across living, non-living and artefactual structures. His understanding of the eye presupposed capacities for the reception of Aristotelian *forms* that could only occur in an animate structure (Meyering 1989, p.45-48).

The shift to taking the living eye, the eye of a dead animal, and an inert device, to be all relevantly similar to one another was a significant one in the history of science. Recognising that the comparison was not always obvious helps us to make note of what is *dissimilar* about these systems: the living eye is highly mobile, and scientists now think that this is essential to understanding how vision works (Findlay and Gilchrist, 2003). Moreover, the habit of thinking of the eye as camera-like hampered scientists' efforts to understand how movement is integral to biological vision.

The moral of this story is that every useful analogy brings with it disanalogies which can mask important features of the target of investigation. Yet analogies like these are at the core of the models which scientists build in order to explain natural systems and mimic their functions. While it is misleading to say that the eye works *just like* a camera, or that both perform the same function of "seeing", the two systems illuminate each other. Building cameras has helped people better understand optics of eye, and knowledge of the eye inspired the invention of photographic cameras.

Computational models of vision are today's version of the camera model of the eye. And as I will show in this chapter, there has been a fruitful exchange between neuroscientists, psychologists, and computer scientists who study vision. The further question, to be considered at the end, is whether this work gives reason to think that vision is *fundamentally* computational. Before then I will say something about what the computational approach to vision is, how it emerged historically, and how we should think about explanatory practice in this area of science.

2. The Computational Approach to Vision

2.1 What Computational Vision Is

Vision is a mental function that has been fairly successfully replicated in artificial intelligence. "Computational vision" often means visual AI or "machine vision". Beyond this, there are the fields of psychology and neuroscience which produce computational models of organic visual systems (Frisby and Stone 2010). The visual system has been the point of origin and/or testing ground for many foundational concepts in computational neuroscience, such as the *receptive field*, *divisive normalisation*, and *hierarchical processing*. Because vision is the most well-studied of all of the perceptual modalities, the computational theory of vision is one of the most advanced parts of the computational theory of mind.

This pre-eminence is the outcome of focussed research since the 1960's on the combined challenges of reverse engineering animal visual systems, creating theoretical frameworks which specify the problem of vision, and developing adequate computational solutions. David Marr's book *Vision,* published posthumously in 1982, played a pivotal role in the story. It synthesised insights garnered over the previous two decades and presented a mathematical theory of visual function at each stage of the system from retinal input to object recognition. It also set an agenda for the subsequent three decades' research in vision science, not least because of the shortcomings of Marr's blueprints for artificial visual systems.

2.2 Neurons and Seeing, Feathers and Flying

The popular account of Marr is that he successfully decoupled the computational approach from the neuroscience of vision, asserting the primacy of computational theory. In this context, the following passage from Marr is frequently quoted:

> "trying to understand perception by studying only neurons is like trying to understand bird flight by studying only feathers: It just cannot be done. In order to understand bird flight, we have to understand aerodynamics; only then do the structure of feathers and the different shapes of birds' wings make sense."
> (Marr 1982, p.27)

The popular account needs to be revised. First, it is worth laying out its essentials. The starting point is Marr's *three levels of explanation*. Marr distinguished *computational, representational/algorithmic,* and *implementational* questions. The first task for the theory of vision is to understand seeing, in most abstract terms, as a computational problem. The second is to show how this problem can be solved using algorithms to transform incoming visual information (e.g. the 2D pixel array at the retina) into different representational formats (e.g. an edge array; a 3D model of objects in a scene). The implementation question – how a physical system like the brain actually runs the algorithm in order to solve the computational problem – is last, and seems trivial compared to the others. On this version of Marr, the conceptual heavy-lifting of vision science is done by computational theoreticians, not experimental neuroscientists: scientific insight flows downwards from the heights of theory to the laboratory bench.

Thus it can be surprising to read that that *primal sketch*, Marr's proposal for the post-retinal stage of representation, was "inspired by findings about mammalian visual systems" (Marr and Hildreth, 1980 p.188), and that Marr and his collaborators were actively working on computational models of simple cells in the early visual system (e.g. Marr and Ullman, 1981). In practice, Marr's research group at MIT pursued a many-pronged strategy which aimed to make progress at each level of explanation and attempted to use knowledge at any one level to constrain the problem space of the others. As Nishihara (2012, p.1017) relates, "Marr wanted … to ground the study of AI in testable theories of the brain's function and architecture."

2. Levels of Explanation versus Levels of Being

On the popular account, one often sees each of Marr's levels being associated with a particular kind of object of investigation, which has its own proprietary discipline: brains, neurons and silicon chips for the implementational level; psychology and software engineering dealing with the representations and algorithms of the middle level; mathematics delivering the overarching computational theory. This way of carving out the levels meshes with the Fodor–Putnam tradition in the philosophy of mind. As an alternative to a reductive physicalism which identifies mental states with brain states, the non-reductive physicalism of this school starts with the notion that mental states can be multiply realized in systems which are physically completely different. Thus psychology (the science of mental representations and cognitive algorithms) will not undergo reduction to any science which deals just with the biological basis of mind (Fodor 1974). Furthermore, because of the many-to-one relationship between neural state and cognitive function, psychology is understood as an autonomous discipline which is unconstrained by findings in neuroscience. These observations about the relations between the sciences of psychology and neuroscience also reinforce the functionalist thesis that the mind, though physical, is irreducible to the brain.

While I accept that the Marrian approach is a non-reductive one, I contend that it is misleading to assimilate it to the functionalist program in the ontology of mind (even if aspects of Marr's programme can be used in support of philosophical functionalism). My central point is that Marr's different levels of explanation are intended to be utilized in the study of *one* concrete system -- the visual brain. Thus, the computational and algorithmic approaches are not proprietary to computer science and psychology,[1] but can also be different modes of explanation *within* neuroscience. This is, in effect, the position taken in theoretical neuroscience, when computational explanations are sought for the presence of specific features of neural systems, such as receptive field structures.[2]

The neurophysiologist Horace Barlow is an influential figure in this tradition of neuroscience, and it is interesting that in a paper which offers a computational explanation for the existence of inhibitory circuitry in the retina (lateral inhibition), he begins with a comparison with the case of flight, anticipating Marr's well known analogy:

> "A wing would be a most mystifying structure if one did not know that birds flew. . . [W]ithout understanding something of the principles of flight, a more detailed examination of the wing itself would probably be unrewarding. I think that we may be at an analogous point in our understanding of the sensory side of the central nervous system. We have got our first batch of facts from the anatomical, neurophysiological, and psychophysical study of sensation and perception, and now

[1] It might be asserted here that the computational approach is proprietary to computer science just because tools and ideas originally developed in that fields were later employed to understand the brain. In response I should point out that by "proprietary" I do not mean just to refer to the disciplinary source of the model. Many models and formal techniques used in special sciences originally came from elsewhere. People would not say that use of an optimality model from economics within evolutionary biology means that the biologist is 'doing economics'. Likewise, they should not think of a neuroscientist using an information-processing model as doing computer science or psychology.

[2] Elsewhere I refer to this as *efficient coding explanation*, which I contrast with *mechanistic explanation* (Chirimuuta 2014). See Sterling and Laughlin (2015) for an extensive (and opinionated) discussion of this approach.

> we need ideas about what operations are performed by the various structures we have examined. . . .
>
> It seems to me vitally important to have in mind possible answers to this question when investigating these structures, for if one does not one will get lost in a mass of irrelevant detail and fail to make the crucial observations." (Barlow 1961, p. 217).

Barlow's point is that it pays to ask, "why?" questions of biological systems, concerning their function and adaptive value, and that these questions should be pursued amidst investigation into structures and mechanisms.[3]

We should think of Marr's computational level questions as a sub-species of the biologist's functional questions.[4] This observation that computational level questions can be raised in the investigation of neural systems themselves makes sense of the continued efforts of Marr and his colleagues within theoretical neuroscience. In contrast, the popular account, which employs a functionalist philosophy of mind and interprets the levels ontologically, precludes computational explanations of the brain itself. The point is just that the functionalist's association of the computational level with mental states and capacities abstracted away from neural hardware makes it hard to see how computational

[3] It is worth making the comparison with Ernst Mayr's (1961) advocacy of non-reductive research strategies in biology, and emphasis on functional "why?" questions. Less well known is the neurologist Francis Walshe's (1961:131) argument that progress in understanding the nervous system will only be made if reductionist strategies are employed in tandem with ones which foreground the function of the system:

> "If we subject a clock to minute analysis by the methods of physics and chemistry, we shall learn a great deal about its constituents, but we shall not discover its operational principles, that is, what makes these constituents function as a clock. Physics and chemistry are not competent to answer questions of this order, which are an engineer's task"

I am not suggesting that it is any more than a coincidence that these three papers were published in 1961, but it does indicate that such ideas were "in the air" at the time. Note that Marr was a doctoral student in Horace Barlow's department, the Physiological Laboratory at Cambridge, before moving to MIT in 1977.

[4] While Shagrir (2010) argues for the importance of "why?" questions in computational level explanations, he does not associate these with functional questions in biology more generally. Thus I go further than Shagrir (2010) in my proposal that we consider Marr and Barlow's theoretical work as part of a wide tradition of functional thinking in biology. See Chirimuuta (forthcoming) for a more detailed presentation of this view.

notions would be relevant to the brain, as opposed to the informational or mental states realized in it. My preferred account also makes sense of the uptake of Marr's distinction between levels of explanation within visual neuroscience today (Carandini 2012). As Frisby & Stone (2012:1042) write,

> "A theory of vision should do more than identify a mechanism that could implement a given computation: it should also provide a functional (e.g. computational) reason why the computation was desirable in the first place. To some extent, Marr's call for a computational account of vision has been taken up in the nascent field of computational neuroscience, where fine-grained analysis of physiological data is commonly interpreted in terms of its functional significance."

In understanding the relationship between the different explanatory levels it is useful to import the concept of *perspectivism* from the philosophy of science. Theoretical frameworks (e.g. Newton's laws, Schrödinger's equation) define "perspectives" (i.e. classical mechanics, quantum mechanics) which are employed to develop models that are then tested against empirical data. As Giere (2006:15) explains,

> "general principles by themselves make no claims about the world, but more specific *models* constructed in accordance with the principles can be used to make claims about specific aspects of the world. And these claims can be tested against various instrumental perspectives. Nevertheless, all theoretical claims remain perspectival in that they apply only to aspects of the world and then, in part *because* they apply only to some aspects of the world, never with complete precision."[5]

Now in vision science – unlike physics, but like other parts of biology – theoretical frameworks are not usually defined by laws of nature; both working models and descriptions of mechanisms have a significant role to play (Machamer et al. 2000). I

[5] Cf. Griffiths et al. (2012, p. 416) in their defense of Bayesian cognitive science: "A theoretical framework provides a general perspective and a set of tools for making models. …. Models are falsifiable, but frameworks are typically not."

suggest that we think of each of Marr's explanatory levels as characterizing kinds of perspectives that scientists use to approach the visual system. The implementational level aligns with a mechanistic perspective in which the goal is to develop detailed and realistic, causal descriptions of the visual brain. Bayesianism can be understood as a representational/algorithmic perspective. Given a clear computational definition of the task of vision as decision making under uncertainty, Bayesianism gives instructions as to how visual information is to be represented and processed in order to yield optimal decisions.

It is helpful to think back to our opening parable of the eye and the camera, and how choices about the relevance of analogies can shape the direction of scientific research. The models and descriptions characteristic of any one explanatory level differ as to which comparisons they take to be relevant to understanding the target system. The computational and algorithmic perspectives invite comparisons between the biological visual system and a man-made computer which is said to perform the same function. Mechanistic perspectives, which look in detail at the nuts and bolts of the system, do not prompt such comparisons because from the point of view of physical implementation the biological and artefactual systems have little in common. No perspective, by itself, can give a complete view of the target system. Ideally the different perspectives should complement each other to give a richer explanation. The mutual interdependence of work at different levels is a note commonly struck in vision science. As Ullman writes,

> "The study of the computations performed by the human visual system can …. lead to new insights and to the development of new and better methods for analyzing visual information. At the same time, given the enormous complexity of the human visual system, a better theoretical understanding of the computations underlying the processing of visual information can supply useful guidelines for empirical studies of the biological mechanisms subserving visual perception."[6]

[6] Quotation from website http://www.wisdom.weizmann.ac.il/~shimon/research.html, accessed 14/10/2016. See also Quinlan (2012:1010), Warren (2012:1054), Frisby and Stone (2012:1043-44). Poggio (2012:1017).

3. Marr's *Vision* Realised?

Because of his tragically early death, David Marr might be thought of as prophet who saw the promised land but never arrived there. He presented a vision of biologically-inspired artificial systems which mimic the tasks which animals perform effortlessly using sight – avoiding obstacles while navigating through a cluttered environment, picking out useful items in a crowded scene. To an impressive degree, these problems have now been solved, though not in the ways that Marr thought they would be. Computationalism, of one kind or another, is now the dominant approach in vision research.[7] In this section I will discuss the merits of deep learning and Bayesian models. What these two strategies have in common is a basically statistical approach to the problem of vision. Through experience of masses of visual information, these models must learn whatever patterns there are in the dataset that facilitate reliable object recognition, for example. The solutions arrived at by statistical methods remain implicit – embedded in the weights of a connectionist network, or the probability distributions of a Bayesian model. In contrast, Marr's programme (like others in the classical "GOFAI" tradition) sought elegant, formal solutions to the problem, ones that could be written down as explicit instructions, such as those for analyzing a complex 3D object as a set of primitive solids.[8]

3.1 Deep Learning and Seeing Machines

[7] An exception is the *empirical paradigm* (Purves et al. 2015). See Orlandi (2014) for philosophical discussion.

[8] See Warren (2012:1055) on the limitations of Marr's programme in comparison with statistical approaches. The "noisiness" of input image data, and neural systems, is another important factor here. Warren (2012:1057) writes that "Marr himself had little enthusiasm for probabilistic approaches because he felt that statistical machinery was not a substitute for the analysis of constraints and information that provides a firm basis for computational theory."

According to one analyst, there will be 10 million self driving cars on the road by the year 2020 (Greenough 2016). If this forecast is to be believed, AI will be driving the first major revolution in land transportation technology since the invention of the automobile. Artificial visual systems are obviously essential to the success of these vehicles. The "deep learning" AI (Talbot 2015) which enables autonomous cars to see roads, lane markings, other cars and pedestrians, is the descendent of neurally inspired network models of the early 1980's.[9]

Fukushima's (1980) "Neocognitron" was inspired by David Hubel and Torsten Wiesel's feedforward, hierarchical description of the cat's visual cortex. The Neocognitron is a connectionist network in which units in different layers are said to perform computations analogous to the simple and complex cells of the brain. The model can be trained to perform pattern classification, for example, recognising hand written numerals. In turn, the Neocognitron inspired other models such as the H-max model (Riesenhuber and Poggio, 1999), and LeNet (LeCun et al., 1995).

The "AlexNet" (Krizhevsky et al. 2012) is considered a landmark development in that it achieves human-level proficiency in the classification of objects presented in complex, real-world images.[10] Following training, the response profiles of the model's units in lower layers resemble the simple bar-like receptive fields of cells in the primary visual cortex. Similarly, Yamins et al. (2014) report that the units in the higher layers of their object recognition network have responses that accurately predict the responses of real neurons in the temporal cortex of the ventral stream.[11]

[9] "Deep learning" refers to AI using artificial neural networks with very many hidden layers. See LeCun et al (2015) for a useful review of kinds of networks and Garson and Bruckner in this volume.

[10] The AlexNet was trained to associate 1.2 million human labelled images with 1000 different object classes. It was tested with 150,000 images and achieved a 37.5% top-1 error rate (network's estimate of most probable classification is incorrect) and 17.0% top-5 error rate (the correct answer does not appear amongst the network's selection of five most probable classification). Networks are now reported to outperform humans in object recognition. See VanRullen (2017) for review and discussion.

[11] The ventral stream is the portion of the visual system thought to be responsible for object recognition, and the temporal cortex is the end stage of the ventral pathway.

Yamins and diCarlo (2016:356) argue that deep learning networks, trained to perform specific ecologically relevant tasks (such as object recognition) are one of the most important tools for understanding the sensory cortex. Their results do indeed indicate that there are intriguing, albeit broad, similarities between the solutions arrived at in nature, and those achieved in computational simulations – most obviously, the repetition of a simple "canonical" computation over many layers of a hierarchical system. But some caveats must be noted. Nguyen et al (2015) showed that deep learning networks are, in their words, "easily fooled", giving high confidence classification responses to patterns which, to a human, look nothing like the actual object. For example, a panel of black and yellow horizontal stripes is confidently classified as an American school bus. Another intriguing weakness of these systems is revealed by "adversarial images". These are images with a few alterations in pixel values, such that they look like normal photographs to a human viewer but lead to gross errors in the artificial network's classification of them. While adversarial images may lead to real security vulnerabilities as such technologies are rolled out in day to day settings (Kurakin et al. 2016), they also raise the question of whether artificial networks are solving the problem of image classification in profoundly non-biological way.[12]

This is the inference made by Piekniewski et al (2016), who also argue that the machine vision community faces a version of Moravec's paradox. Moravec (1988) wrote that,

> "It is comparatively easy to make computers exhibit adult level performance on intelligence tests or playing checkers, and difficult or impossible to give them the skills of a one-year-old when it comes to perception and mobility."

In the case of machine vision, AlexNet has achieved adult-level performance in a task (i.e. the categorisation of thousands of different kinds of objects) which would be beyond

[12] But see Yamins and Di Carlo (2016:363). They speculate that it would be possible to create adversarial images to trick a human visual system if the individual's detailed neural circuitry were known.

a human infant, or presumably any non-human animal. However, it has not yet been shown that artificial systems can solve basic sensory-motor coordination problems (like following a moving object in a complex, inpredictable environment) which year-old babies and all other animals do reliably.

Piekniewski et al (2016) argue that a way forward is to make artificial networks more biologically realistic, by including feedback connections (which outnumber feedforward pathways in the sensory cortex) and lateral connectivity within network layers, so that both temporal and spatial context influence the computations in any one unit. Another more naturalistic feature of their work is that they train their network not with photographs but with videos, thus mimicking the continuous stream of sensory input that falls on the eyes. Their model does show good performance in tracking a single object, such as a green ball, recognizing its location in the frame over a variety of background colours, textures and lighting conditions.

3.2 Bayesian Brains and Optimal Solutions

Bayesianism is now one of the central approaches in the modelling of perceptual systems. It is a compelling manifestation of the inferentialist idea that perception is the process of drawing conclusions about states of affairs in the external world on the basis of limited sensory data which underdetermine their distal cause. While Marr emphasised the impoverishment of the external world information captured by the 2D projections onto the retina, Bayesians begin with the point that the data received by the visual system is noisy and ambiguous in a variety of ways, not least because of the stochastic nature of sensory transduction and neural transmission. Bayesian inference can provide optimal strategies for decision-making under uncertainty. So it follows that it is at least theoretically interesting to compare the predictions of Bayesian models of the visual systems with the outputs of their biological counterparts.

While Bayesian models have come to the foreground really in the last twenty years, the essence and motivation for the approach was stated by Horace Barlow some decades before:

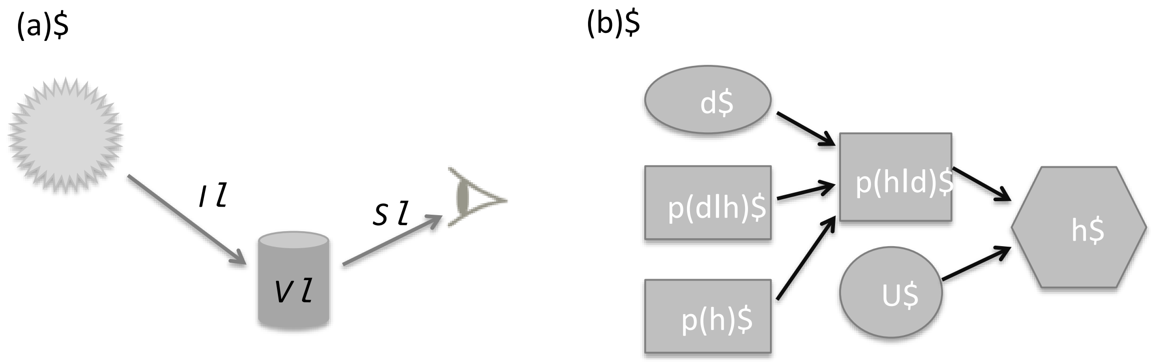
> "Certain sensory events have occurred, and what one optimally needs are the probabilities of these events on certain hypotheses. For instance, if I enter my darkened home and come to the point where there is a swing door, I must decide whether to put out my hand and push it open or whether to walk straight on. The decision should be made partly on the basis of what I can see, but this information must be combined with the prior probabilities of the door being open or closed, and also the payoffs and penalties of the various outcomes; the survival value of walking through a closed door is obviously lower than it would be for the other possibilities. However, these are not sensory problems; what we need from our eyes is simply the probability of the sensory events occurring if the door is closed, and so on."
> (Barlow 1969:878)

The modelling of colour constancy gives a useful example of the Bayesian approach.[13] Colour vision is a useful guide to our surroundings because the colour appearances of objects are relatively stable across changes in illumination, such that within a very broad range of lighting conditions at the breakfast table, we can tell whether our toast is burned or not, or if our juice is apple or cranberry, just by looking at the colours. However, the challenge of achieving constant colour perception is set by the physics of light reflection and absorption. The wavelength of light reaching the eye after being reflected from the glass of juice is due both to the absorption–transmission profile of the liquid and the wavelength of incident light – which is quite different if coming from the sun through the window behind the table, or from an overhead bulb on a cloudy morning (see Figure 1a). On the Bayesian approach, colour constancy is achieved when the visual system makes a correct inference about the reflectance profile of the object viewed, disambiguating it

[13] See Allred (2012) for more of the scientific details discussed here and some useful reflections on different modelling strategies in vision science. See Kersten et al (2004) on the Bayesian approach to object perception, and Rescorla (2015) for a philosophy-oriented overview.

from the spectrum of the illumination. This is especially challenging because information loss occurs due to the broad, non-specific spectral sensitivity of the cones in the retina.

FIGURE 1



(a) *The Challenge of Colour Constancy.* The wavelength of the light incident on the eye ($S\lambda$)is a product of the wavelength of the ambient illuminating light ($I\lambda$) and the spectral absorption-transmission function of the glass of juice ($V\lambda$) Thus the proximal stimulus at the eye confounds the spectral properties of the perceived object and the illuminant. Colour constancy is achieved if the visual system is able to disambiguate these two physical sources of colour changes.

(b) *Schematic of Bayesian Inference*. The visual system receives uncertain sensory data (*d*) regarding conditions in the external world. In order to arrive at a correct hypothesis about external conditions (*h*), the optimal strategy is to calculate the posterior distribution, $p(h/d)$, from the likelihood distribution, $p(d/h)$, and prior $p(h)$, according to Bayes' theorem. The final hypothesis is calculated by weighting the posterior with a utility function, *U*.

Figure 1b is a schematic of the Bayesian approach. The output of the Bayesian model is a guess (*hypothesis)* about what state of the world caused the sensory data. In the case of a colour constancy model, the final hypothesis is a guess as to the reflectance profile of a coloured object. Brainard and colleagues (2006) compared the output of their model with empirical data on the constancy of a simple array of tiles depicted under different illuminants and a good fit was achieved, though the authors note that the experimental conditions were not chosen specifically to test the model (p.1277). A particularly intriguing comparison is that the model correctly predicts errors in colour constancy judgments when the illumination is purple (Allred 2012 p.221). Because purple is an improbable colour for lighting (has low prior probability) the Bayesian algorithm is biased towards inferring that the purple hue is a property of the object, not the illuminant. It would seem that the human visual system is biased in the same way.

Bayesianism provides vision science with a powerful theoretical framework for understanding the computational task of vision, as well as flexible modelling tools. Its use has not been restricted to human visual capacities with, for example, Schloltyssek et al. (2016) reporting that colour perception in chicks can be described with a Bayesian model. Bayesian models have also been employed to explain individual and population responses of neurons, thus giving insights into how neurons might perform statistical computations (e.g. Beck et al. 2008). This holds out the promise that the Bayesianism will be a comprehensive theoretical framework for investigation of both the psychophysics and neurophysiology of sensory perception.

Assuming, for the sake of argument, that the prophets of Bayesianism are right, and that the approach is unmatched in its scope and explanatory power, does this mean that the visual brain *is* Bayesian, that seeing *is* Bayesian inference?[14] Rescorla (2015) urges us to take a realist stance here, bolstered by the explanatory successes of Bayesian models. The claim is that the visual system approximately implements Bayesian inference, and that its inner states approximate to the representations of the probability distributions which feature in the models. As an alternative to realism, Rescorla only considers the

[14] While the "prophet of Bayesianism" is intended to be a caricature, Rescorla (2015) argues something along these lines. And see Knill and Pouget (2004:713) who infer the "Bayesian coding hypothesis" (that the brain houses probabilistic representations of sensory information) from the "Bayes optimality" of human performance in certain psychophysical tasks. Note that the Bayesian coding hypothesis is less contentious -- and more specific -- than the claim that biological perception is Bayesian inference.

instrumentalism of Colombo and Seriès (2012), where the models are interpreted as "tools for predicting, systematizing and classifying statements about people's observable performance" (p.705).[15]

As an alternative both to realism and instrumentalism, I contend that we should be *perspectival realists* (Giere 2006) and pluralists when interpreting Bayesian and other families of models in vision science. The idealised nature of Bayesian models – the fact that they make assumptions known to be false of biological systems, and posit intractable computations – is a fly in the ointment for realism. Rescorla (2015) bets on a de-idealizing trajectory of future research, where models eventually become more realistic descriptions of actual processes.

This optimism neglects the broader issues about the role of idealized models in science. It is instructive to compare Bayesian models with other optimality models used in biology. In brief, optimality models examine, in most general terms, the challenges and constraints imposed upon a biological system, and derive the theoretically optimal strategy for dealing with this situation. They are often highly abstract, describing none of the specifics of the concrete situation, and idealized in the sense of making assumptions known to be false (e.g. that there is an infinite breeding population). Batterman and Rice (2014) discuss Fisher's sex-ratio model in such terms, as a "minimal model" and a "caricature" of nature. Their idea is that the optimality model is applicable to natural populations because it stands in the same "universality class" of systems whose behaviour converges on distinctive patterns of behaviour. The systems within the same universality class are still fundamentally different in many other respects, most obviously physical constitution, but also regarding dispositions to perform other kinds of functions.

Now Bayesian models describe optimal solutions for perceptual decision making under uncertainty.[16] Real systems, to the extent that their behaviour is accurately predicted by

[15] Cf Orlandi (2014:88) "although usefully described by a Bayesian framework, the visual system need not perform Bayesian inferences in any substantive sense. The system proceeds in rough accord with Bayesian norms….. but this just means that the system acts *as if* it is a Bayesian observer."

the model, fall within the same universality class as the model. But nothing within the Bayesian approach assumes that biological perceivers are always optimal, or that there are not radical differences between the strategies employed in nature, and described in the model. As Knill and Pouget (2004, p.718) point out, departures from optimality are an important focus for research, and we know in advance that in a strict sense the brain is *not* Bayesian:

> "an equally important challenge for future work is to find the ways in which human observers are not optimal. Owing to the complexity of the tasks, unconstrained Bayesian inference is not a viable solution for computation in the brain."

As I have argued elsewhere (Chirimuuta 2017), the comparison between a neural system and a theoretical solution offers a *mathematical explanation* of why the system evolved to have the computational properties which are observed – for example, why sensory systems weight the impact of incoming data according to its uncertainty, or why neural populations might employ probabilistic representation. The mathematical theory of decision making gives a priori reason to think that Bayesian processing is relevant to biology, but it is just one perspective on a complex biological system.[17] This is how Griffiths et al. (2012:421) make the point:

> "Different theoretical frameworks, such as Bayesian modeling, connectionism, and production systems, have different insights to offer about human cognition, distributed across different levels of analysis. A connectionist model and a Bayesian model of the same phenomenon can both provide valuable information—one about how the brain might solve a problem, the other about why this solution makes sense—and both could well be valid."

16 In perceptual science, the optimal solution is usually cashed out as the answer given by an "ideal observer"—a hypothetical agent who extracts the most information ("signal") from the noisy data provided by the sensory system.

17 Cf. Frisby and Stone (2012:1047): "This is not intended to imply that Bayes's theorem is a dominant feature of visual processing, but that, by virtue of being the only rational basis for inference under uncertainty, it is a necessary feature of visual information processing."

Thus the corollary of perspectival realism is *pluralism* – the thesis that a number of different theoretical frameworks and modelling strategies should be employed to build a more well-rounded understanding of the natural system.[18]

To conclude, we can think again of the eye and the camera. The photographic camera, a man-made system with none of the complexity or evolutionary heritage of the eye does still instantiate the same optical laws which explain image formation on the retina. Likewise, simple and highly idealised models, in various theoretical traditions, capture some of the computational principles which are employed by the visual system and explain its capacities. Without doubt there is much still to be discovered, and the comparison between the visual brain and a computational model might come to be seen to future generations both as compelling and naïve as the camera model of the eye seems to us now.[19]

REFERENCES

Allred, S. (2012) Approaching color with Bayesian algorithms. In G. Hatfield and S. Allred (eds.) *Visual Experiences: Sensation, Cognition, and Constancy.* Oxford: Oxford University Press.

[18] Another corollary is that the accuracy of models is only evaluable from *within* as perspective, not across perspectives (Giere 2006). So I think perspectival realism, and not instrumentalism, gives the better account of this statement from Griffiths et al. (2012:421):

> "The ultimate test of these different theoretical frameworks will be not whether they are true or false, but whether they are useful in leading us to new ideas about the mind and brain, and we believe that the Bayesian approach has already proven fruitful in this regard."

[19] I am much indebted to Joe Howse and to the editors of this volume for thoughtful and helpful comments on this chapter.

Barlow, H. B. (1961). Possible principles underlying the transformation of sensory messages. In W. A. Rosenblith (Ed.), *Sensory Communication*. Cambridge, MA: MIT Press.

Barlow, H. B. (1969) Pattern recognition and the responses of sensory neurones *Ann. Acad. Sci.* **156** 872–81

Batterman, R. and C. Rice (2014). Minimal model explanations. *Philosophy of Science 81*(3), 349–376.

Beck, J. M., W. J. Ma, R. Kiani, T. Hanks, A. K. Churchland, J. Roitman, M. N. Shadlen, P. E. Latham and A. Pouget.

Bowers, J. and Davis, C. 2012: Bayesian just-so stories in psychology and neuroscience. *Psychological Bulletin*, 138, 389–414.

Brainard, D. H., P. Longère, P. B. Delahunt, W. T. Freeman, J. M. Kraft and B. Xiao. (2006). "Bayesian model of human color constancy." *Journal of Vision* 6, 1267–1281

Carandini, M. (2012). From circuits to behavior: A bridge too far? *Nature*, *15*(4), 507–509

Chirimuuta, M. (2014). Minimal models and canonical neural compu- tations: the distinctness of computational explanation in neuroscience. *Synthese 191*, 127–153.

Chirimuuta, M. (2017) Explanation in neuroscience: causal and non-causal. *British Journal for the Philosophy of Science*. https://doi.org/10.1093/bjps/axw034

Chirimuuta, M. (forthcoming). "Marr, Mayr, and MR: What Functionalism Should Now Be About". *Philosophical Psychology*

Colombo, M. and Seriès, P. 2012: Bayes in the brain—on Bayesian modeling in neuroscience. *British Journal for the Philosophy of Science*, 63, 697–723.

Findlay, J., and I. Gilchrist. 2003. *Active Vision: The Psychology of Looking and Seeing*. Oxford: Oxford University Press.

Fodor, J. A. (1974). Special sciences (Or: The disunity of science as a working hypothesis). *Synthese*, *28*, 97–115.

Frisby J P, Stone J V, 2010 *Seeing*: *The Computation Approach to Biological Vision* (Cambridge, MA: MIT Press)

Frisby, J P and J V Stone (2012) "Marr: An Appreciation" *Perception*. 41: 1040-1052

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition una↵ected by shift in position. *Biological cybernetics*, 36(4):193–202.

Giere, R. N. 2006. *Scientific Perspectivism*. Chicago: University of Chicago Press

Greenough, J. (2016) "10 million self driving cars will be on the road by 2020". *Business Insider*. http://www.businessinsider.com/report-10-million-self-driving-cars-will-be-on-the-road-by-2020-2015-5-6 Accessed 14/10/2016.

Griffiths, T. L., N. C. Chater, D. N. Norris and A. Pouget (2012) "How the Bayesians Got Their Beliefs (and What Those Beliefs Actually Are): Comment on Bowers and Davis" *Psychological Bulletin* 138(3): 415–422

Kersten, D., P. Mamassian and A. Yuille (2004). "Object perception as Bayesian inference." *Annu. Rev. Psychol.* 55:271–304

Knill, D. C. and Pouget, A. [2004]: ‘The Bayesian Brain: The Role of Uncertainty in Neural Coding and Computation’, Trends in Neurosciences, 27, pp. 712–9.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.

Kurakin, A., Goodfellow, I., and Bengio, S. (2016). Adversarial examples in the physical world. *arXiv preprint. arXiv:1607.02533*. https://arxiv.org/pdf/1607.02533v2.pdf

LeCun, Y., Jackel, L., Bottou, L., Brunot, A., Cortes, C., Denker, J., Drucker, H., Guyon, I., Muller, U., Sackinger, E., et al. (1995). Comparison of learning algorithms for handwritten digit recognition. In *International conference on artificial neural networks*, volume 60, pages 53–60.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444

Machamer, P., L. Darden, and C. F. Craver (2000). Thinking about mechanisms. *Philosophy of Science 67*, 1–25.

Marr, D. 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: Freeman.

Marr, D. and S. Ullman (1981). “Directional sensitivity and its use in early visual processing.” *Proc. R. Soc. Lond.* B 211, 151-180.

Mayr, Ernst: 1961, 'Cause and Effect in Biology', *Science* 134, 1501-1506.

Meyering, T. C. (1989). *The Historical Roots of Cognitive Science.* Dordrecht: Kluwer.

Moravec, H. (1988). *Mind children*, volume 375. Cambridge Univ Press.

Nishihara, H K (2012) “Recollections of David Marr” *Perception* 41: 1027-1030

Nguyen, A., Yosinski, J., and Clune, J. (2015). Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. *In 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 427–436. IEEE.

Orlandi, N. (2014) *The Innocent Eye.* Oxford: Oxford University Press.

Piekniewski, F., P. Laurent, C. Petre, M. Richert, D. Fisher and T. L. Hylton (2016). “Unsupervised Learning from Continuous Video in a Scalable Predictive Recurrent Network” *arXiv preprint. arXiv:1607.06854. https://arxiv.org/pdf/1607.06854v3.pdf*

Poggio. T. (2012) “The *Levels of Understanding* Framework, Revised” *Perception* 41: 1017-1023

Purves, D., Y. Morgenstern and W. T. Wojtach (2015) “Perception and Reality: Why a Wholly Empirical Paradigm is Needed to Understand Vision” *Front. Syst. Neurosci.*, 18 November 2015 http://dx.doi.org/10.3389/fnsys.2015.00156

Quinlan, P. (2012). “Marr's *Vision* 30 Years on: From a Personal Point of View” *Perception* 41: 1009-1012

Rescorla, M. (2015) “Bayesian Perceptual Psychology” in M. Matthen (ed). *The Oxford Handbook of Philosophy of Perception.* Oxford: Oxford University Press.

Riesenhuber, M. and Poggio, T. (1999). Hierarchical models of object recognition in cortex. Nature neuroscience, 2(11):1019–1025.

Schloltyssek, C., D. C. Osorio and R. J. Baddeley (2016). Color generalization across hue and saturation in chicks described by a simple (Bayesian) model. *Journal of Vision*. 16(8). doi:10.1167/16.10.8

Shagrir, O. (2010). "Marr on computational-level theories." *Philosophy of Science.* 77(4):477-500.

Sterling, P. and S. B. Laughlin (2015). *Principles of Neural Design*. Cambridge, MA: MIT Press.

Talbot, D. (2015) "CES 2015: Nvidia Demos a Car Computer Trained with 'Deep Learning'" *MIT Technology Review*. https://www.technologyreview.com/s/533936/ces-2015-nvidia-demos-a-car-computer-trained-with-deep-learning/ Accessed 14/10/2016.

vanRullen, R. (2017) "Perception Science in the Age of Deep Neural Networks" *Frontiers in Psychology.* Vol. 8, Article 142. doi: 10.3389/fpsyg.2017.00142

Walshe, F. M. R. (1961) "Contributions of John Hughlings Jackson to Neurology: A Brief Introduction to His Teachings" *Archives of Neurology* 5:119-131

Warren, W. H. (2012). "Does This Computational Theory Solve the Right Problem? Marr, Gibson, and the Goal of Vision" *Perception* 41: 1053-1060

Yamins, D.L.K., H. Hong, C. F. Cadieu, E. A. Solomon, D. Seibert and J. J. DiCarlo (2014) Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc. Natl. Acad. Sci. USA* **111**, 8619–8624

Yamins, D. L. K and J. J. DiCarlo "Using goal-driven deep learning models to understand sensory cortex". *Nature Neuroscience.* 19(3):356-365.