Can Computational Goals Inform Theories of Vision?

Barton L. Anderson

School of Psychology, University of Sydney

Abstract

One of the most lasting contributions of Marr's posthumous book is his articulation of the different "levels of analysis" that are needed to understand vision. Although a variety of work has examined how these different levels are related, there is comparatively little examination of the assumptions on which his proposed levels rest, or the plausibility of the approach Marr articulated given those assumptions. Marr placed particular significance on computational level theory, which specifies the "goal" of a computation, its appropriateness for solving a particular problem, and the logic by which it can be carried out. The structure of computational level theory is inherently teleological: What the brain does is described in terms of its purpose. I argue that computational level theory, and the reverse-engineering approach it inspires, requires understanding the historical trajectory that gave rise to functional capacities that can be meaningfully attributed with some sense of purpose or goal, that is, a reconstruction of the fitness function on which natural selection acted in shaping our visual abilities. I argue that this reconstruction is required to distinguish abilities shaped by natural selection—"natural tasks" —from evolutionary "by-products" (spandrels, co-optations, and exaptations), rather than merely demonstrating that computational goals can be embedded in a Bayesian model that renders a particular behavior or process rational.

Keywords: Computational theory; Bayesian modeling; Teleology; Natural tasks

1. Introduction

The central tenet of Marr's approach is that vision should be understood as an information-processing task, which he characterized as a sequence of transformations that map one representation onto another. The tri-level hypothesis captured different kinds of questions that could be asked about the nature of these mappings. Computational level questions concern the presumed *goal* or *purpose* of a mapping, that is, the specification of the "task" that a particular computation "solved." Algorithmic level questions involve

---

Correspondence should be sent to Barton L. Anderson, University of Sydney, Sydney, NSW, Australia. E-mail: barton.anderson@sydney.edu.au

specifying how this mapping was achieved computationally, that is, the formal procedure that transforms an input representation into an output representation. Implementation level questions play a relatively incidental role in Marr's approach, being relegated to addressing questions concerning how a particular solution to a problem happens to be implemented physically.

There were a number of immediate consequences of Marr's characterization of vision. First, the emphasis on computational level theory inverted the meaning of what it meant to explain vision. Whereas the prevailing mechanistic view explained vision "*in terms of* neurons" (p. 18),[1] Marr asserted that "the nature of the computations that underlie perception depends more upon the particular problems that have to be solved than upon the particular hardware in which their solutions are implemented" (p. 27). Perception was, in this view, best understood and explained by characterizing the particular "problems that have to be solved." To this end, each transformation in an information-processing sequence is ascribed a particular purpose, which putatively provides the basis for both understanding *what* the visual system computes and *why* it computes it. To Marr, an explanation of vision is explicitly teleological: Knowing what the visual system does and how it accomplishes it plays less of an explanatory role than understanding its goal or purpose.

There are a number of issues that are raised by Marr's approach, many of which will be taken up in other contributions to this issue or have already been discussed elsewhere (see *Perception*, 2012, *41* (9)). To avoid as much redundancy as possible with previous commentaries, I will focus on his characterization of perceptual processes as *solving problems*, *performing tasks*, or having *goals* or *purposes*. Although some authors have questioned whether Marr correctly captured the *particular* tasks the visual system performs, there has been surprisingly little discussion on the teleological structure of his framework, and whether this structure provides any meaningful theoretical or empirical leverage into understanding vision. In what follows, I will focus on two aspects of Marr's approach: (a) the conceptual issues entailed by teleological descriptions of visual processes; and (b) whether such characterizations provide any leverage in identifying the putative tasks visual systems perform or the goals they achieve. In what follows, I will focus on how Marr's ideas were articulated within the domain of vision, and how such ideas have developed or been adapted subsequently in attempting to craft models of visual competence and experience.

2. The purpose of vision

As a computer scientist, or in engineering more broadly, it is natural to begin by specifying a problem that needs to be solved. The problem motivates the general computational strategy and algorithms used to craft a solution. Engineers explicitly design systems to perform particular tasks or serve specific functions. The difficulty arises when attempting to apply the same logic to the design of biological systems. Marr's concept of computational level theory depends critically on the relevance of engineering concepts

for understanding biology and psychology. Indeed, the contention that vision can be ascribed a purpose (or the specification of the "goal" of a computation) expresses a claim that biological processes can be shaped in the *service of* particular ends. The legitimacy of computational level theory rests on this assumption, so it is worth exploring this idea to determine what this claim actually entails, both theoretical and pragmatically (i.e., how or whether such ideas can or cannot inform theory construction or experimental practices).

There is a significant body of literature within biology, psychology, and philosophy about the scientific legitimacy of teleological explanations. It is not possible to provide even a modest survey of this literature within the space constraints of the current manuscript, nor is it the purpose of the present discussion. Rather, my goal is to provide a sufficiently clear understanding of what the idea of a computational level theory entails within the context of the evolution of the visual system to determine what it would mean to engage in computational level theory as a general research strategy within the domain of vision. There are both principled and pragmatic reasons for restricting the following discussion to the role of teleological explanations in the context of evolution. The pragmatic reason is simply a matter of scope. Although teleological explanations can also arise in the context of learning, the literature on learning is vast, and the nature of this series of papers is to provide a collection of circumscribed pieces on the significance of Marr's approach for current topics in cognitive science. The principled reason follows from the fact that learning played essentially no role in Marr's original representational framework for vision, and hence cannot provide any insight into meaning of vision's "purpose" in the framework in which he originally developed his ideas. This implies that the validity of such constructs in his original work must rest in the sense of "engineering" embodied in evolutionary theory. It is, therefore, instructive to explore the sense in which evolutionary processes can be construed as generating a visual system that can be described as embodying a purpose or achieving goals, and how (or whether) such processes provide any theoretical or empirical leverage into understanding vision.

There are a number of ways to formalize evolutionary processes that could be used to provide insight into what it would mean to shape vision to achieve particular goals (see, e.g., Lewontin, 1996; Rosen, 1985, 2012). Although different formalizations may differ in their details, there are two essential theoretical constructs that play a pivotal role in all such analyses: *fitness* and *selection*. In what follows, I recapitulate Rosen's (1985, 2012) formalization of these concepts as they relate to function and adaptation, which will motivate the subsequent discussion on how these ideas can (or cannot) provide conceptual or empirical leverage into understanding vision.

Consider a simple organism with a simple tropism, such as a negative phototropism. For simplicity, assume that its movements are restricted to a 2D surface. The intensity of light E at each point on this surface has a definite value, which in turn defines a gradient ∇E at each point on the surface. We further assume that the relevant behavior or *phenotype* of the organism is captured by the direction it moves in response to the gradient, which we well denote dx/dt and dy/dt. The organism's direction of motion depends on

both the gradient, ∇E, and by the animal's genome, α, satisfying equations of the form:

$$dx/dt = \varphi(\alpha, \nabla E) \tag{1}$$

$$dy/dt = \psi(\alpha, \nabla E)$$

We further assume that the genetic variation in organisms α_i generates variability in how they move in response to the gradient ∇E. The movement of each organism thereby generates a trajectory or path in our 2D surface, which will vary depending on the value of α in (1) (i.e., different species will generate different paths). Note that the directions of motion constitute the organism's phenotype (behavior) in this example, α is its genome, and ∇E represents animal's observable environment.

Let us then suppose that we have another function, $U = U(x, y)$, which can also be defined on this surface and satisfies the same properties as E such that a gradient ∇U can be defined at each point. We can imagine, for example, that such gradients specify something like a predator density, which defines a vector field of the form:

$$dx/dt = \partial U/\partial x \tag{2}$$

$$dy/dt = \partial U/\partial y$$

Note that there is currently nothing that links the paths in (1) with the integral curves in (2); they simply represent two vector fields defined on the same 2D space that can be used to define paths. We can, however, construct some metric that will compare these two curves that captures how discordant the paths traced in (1) are from (2), such as the areal difference between the paths, say $D(t)$. This function provides a measure of how far the path generated by an organism following Eq. (1) departs from the integral curve of (2) to which it is currently "blind." The inverse of this measure of discordancy, $F(t) = 1/D(t)$, will be defined as the *fitness* of the path, such that one path is considered more *adaptive* than another if its fitness is larger.

To convert this measure of discordancy into something that *links* the paths in (1) and those in (2), something is needed that *selects* different paths (phenotypes) on the basis of their relative fitness (i.e., their relative discordancy). Note that, in the context introduced above, the variability in phenotypes described in (1) is determined solely by the variability in their genotypes; the variation in genotype captures how a given species α_i responds to the gradients in the environment ∇E. The role of selection is to favor those species whose response to the gradient ∇E most closely follows the gradients ∇U. Genomes, therefore, evolve through a selection process that drives the initially unrelated paths in (1) toward the integral curves in (2). The selection of different phenotypes, which vary in their measure of relative fitness, can therefore be expressed as an evolution in the space of genomes. More specifically, the fitness function F can be used to define a space of genomes A, such that each element α in A can be associated with a

fitness $F(\alpha)$. In a manner analogous to (1) and (2), the space can be used to construct a vector field and dynamics on the space of genomes of the general form:

$$d\alpha/d\theta = K\nabla F(\alpha) \tag{3}$$

where θ is a time parameter associated with the generation time of species in A. It is this equation that captures the selection process and "engineering force" of evolution: it expresses how biological systems are shaped via natural selection's response to gradients of fitness toward states that are increasingly adaptive.

There are a number of aspects to the preceding development that are relevant to understanding Marr's concept of computational level theory. First, the teleological ascription of purpose to visual processes emerges as a consequence of two interrelated concepts in the context of evolution: *fitness* and *selection*. The concept of fitness is inherently relational; it captures some measure of the discordancy between a condition of an animal's environment and the animal's behavior; that is, it provides information about the *consequences* of an organism's behavior in a particular environment, by capturing something about an animal's current behavior, and its relationship with an optimal behavior given its environment. In the example developed above, if $U(x, y)$ in (2) corresponds to predator density, and predator density in U is higher in regions where E is higher, then those species that follow the gradients of E will automatically track the gradients in U. Selection links the animal's behavior to its *consequences* by increasing the fecundity of those species α_i whose behavior with respect to E most closely track the gradients in U, driving biological systems toward increasingly higher degrees of fitness.[2]

The preceding articulates a general relationship between evolution, fitness, selection, and adaptation. We have, in this context, only considered a very simple system, but it nonetheless provides a general sense by which vision could be shaped by natural selection to accomplish particular tasks or perform particular functions. The key insight from the preceding development is any imputation of vision's purpose requires reconstructing the dynamics expressed in Eq. 3, that is, *inferring the fitness function that shaped the evolutionary history of our visual system*. In what follows, I consider some of the significant theoretical and pragmatic obstacles that arise in attempting to make such inferences.

3. How do you determine the computational goal(s) of vision?

The preceding section provides some understanding of how evolution could "engineer" a visual system to perform particular tasks or serve particular functions. The upshot of this analysis is that the ascription of computational goals to a species rests on the capacity to infer the fitness function that shaped its visual competencies. The question is: how can this be accomplished? If vision is to be understood as satisfying computational goals, how do you infer the historical fitness function that shaped what those goals are? Although Marr advocated the importance of approaching the analysis of the visual system

by starting with a computational level theory, he provided no insight for how such theories could be constructed or evaluated. Marr simply *guessed* the goal of vision and proceeded on the basis of that guess. He assumed that the primary purpose of vision was to build "a description of the shapes and positions of things from images" (p. 36). He treated other aspects of vision as secondary and presumed that many dimensions of visual experience could be "hung off a theory in which the main job of vision was to derive a representation of shape" (p. 36). If his guess is wrong, then many of the experimental findings about vision at the physiological or psychophysical level will appear mysterious, inexplicable, or simply uninterpretable.[3] Indeed, almost none of Marr's specific ideas have withstood the test of time; his most persisting contribution is a general belief in the utility of computational level theory and the reverse-engineering approach it inspires.

There are a host of issues that emerge when attempting to reconstruct the engineering of a visual system generated by an evolving fitness function. Although the problems confronted by an engineer in designing a system to perform a task are specified at the outset, the "problems" a species comes to solve emerge over the course of its evolution. *Only those properties that were shaped by natural selection on the basis of their consequences can be truly deemed an adaptive "purpose" or "goal" of vision.* The fact that something is currently used for a particular purpose or functions in a particular way is insufficient evidence that it was *selected* for that purpose (also known as the problem of "spandrels"; see Gould & Lewontin, 1979; Lewontin, 1996). Even the dimensionality of the fitness function remains unknown, and almost certainly varies over the course of evolution. Although it is a truism that *some* fitness function shaped our visual system to satisfy particular ends, there is currently no clear means of inferring what those ends are, and hence no clear way to construct a general computational level theory of vision without some understanding of the particular fitness function that led to our current perceptual abilities.

The complexity of generating a general computational level theory has not, however, dissuaded a variety of authors from attempting to craft computational level theories for subdomains of vision science. Although the particular problems vary, the structure of the argument is relatively generic. This argument rests on the claim that our perceptual competencies, abilities, and/or experiences should be conceptualized as solutions to particular computational *recovery* or *estimation* problems (such as shape, color, lightness, size, depth, etc.). The presumption is that the visual system evolved to *estimate* particular objective dimensions of the environment, and that those organisms whose perceptual outputs more closely approximated ground truth would be more adaptive. The general argument takes the form: It is important to recover (or estimate) environmental property X because it is an intrinsic property of the environment. Hence, knowledge of property X should enhance the overall fitness of the animal, which natural selection shapes to be performed (approximately) optimally. This generic argument can be (and typically is) applied to any property of an animal's environment. Presumably, however, some perceptual abilities were more relevant to our survival than others, and therefore contributed more to our overall fitness. But which ones? And how can one decide on the basis of data?

Bayesian ideal observer theory has been forwarded as a possible solution to this problem. Geisler and Diehl (2003) argued that Bayesian ideal observer theory formalizes

Marr's "intuitive" notion of computational level theory, which suggests that it may provide some leverage into the evaluation and/or articulation of computational level theories. Indeed, the Bayesian framework has been applied to the problem of natural selection, which suggests that it could provide some means of distinguishing functions shaped by natural selection from evolutionary by-products. It is, therefore, worth considering whether this framework can provide any insight into either the articulation or evaluation of computational level theories.

Bayesian ideal observer analysis begins by specifying a particular "natural task"; the modifier "natural" presumably refers to the assertion that a particular task being studied was shaped by natural selection to perform that particular function. This task plays the identical role that computational level theory plays in Marr's framework; it specifies a particular problem to be solved. Once a task is specified, an inventory of the environment is conducted to assess what information is available to solve it. Two kinds of information are needed for a Bayesian model: a prior distribution, which lists the probability of all of the different states of the world that the visual system can encode (such as the distribution of objects, materials, shapes, lighting, etc.); and a likelihood distribution, which lists the probability of the stimulus given each possible state of the environment (and hence embodies some "understanding" of the laws that convert states of the world into images, such as the interaction between light and surfaces, laws of perspective projection, and effects of viewpoint).[4] Bayes theorem can then be used to derive the most probable interpretation of a given perceptual input. Perceptual experiments are then performed to assess how human performance compares to the Bayesian ideal. If the chosen task or problem was indeed "natural," that is, a direct product of natural selection, the expectation is that human performance will in some sense approximate ideal performance. In principle, one might think that this process contains the information needed to distinguish those perceptual abilities that were shaped by natural selection from those that were not, and are hence worthy of being dubbed "natural tasks." In practice, however, no such distinctions are typically made (see, however, Anderson, O'Vari, & Barth, 2011). In almost all published cases, some Bayesian model can be found that approximates or parallels human performance. As there is no criterion available to judge whether the discrepancy between human performance and the Bayesian ideal performance is too large for the task to be construed as "natural," the similarity between human and ideal data completes an explanatory circle: It is taken as evidence that the visual system performs something equivalent to Bayesian inference, and that the task being studied is indeed "natural" in the sense of being a product of natural selection.

How, then, are particular tasks chosen as candidate "natural tasks" of vision? Despite the presumed need to begin with a computational goal, in actual practice, computational theory—or the specification of a natural task—typically begins by identifying something that vision *seems to do,* and then re-expresses a perceptual outcome as the solution to a computational problem.[5] Indeed, many of the "natural tasks" that have been evaluated have already existed as domains of inquiry within the field of vision science. Take, for example, our experience of surface lightness. The Bayesian or computational level theorist converts this dimension of experience into the answer to the question: What is a

surface's diffuse reflectance (the presumed physical "source" the visual system evolved to recover)? Once this perceptual experience is posed as the answer to a computational question, it is then shown that there is insufficient information in the images to answer this question uniquely, since the image data contains a conflated mixture of 3D geometry, surface reflectance, and the illumination field. This, in turn, motivates the treatment of perception as a form of probabilistic inference. What began as a dimension of psychological *experience* is transformed into a computational recovery problem based solely on the presumption that the experience in question (perceived lightness) is something that the visual system attempts to derive from images. The perceptual *impression* of lightness then becomes a feat in which the visual system must go beyond the available image data to guess a distal cause, which turns vision into a collection of probabilistic inference problems.

But what justification is there that the visual system was actually shaped by natural selection to compute a quantity like diffuse reflectance? How do we know that the experience of lightness is not the epiphenomenal output of a system "designed" to compute other quantities (i.e., a perceptual "spandrel")? This question is almost never explicitly addressed. Rather, the recovery of any given environmental property is primarily justified by the observation that our experience can, at least in some contexts, be *closer to* its distal cause (diffuse reflectance) than to some proximal image variable from which it is putatively derived (such as luminance). When that veridicality persists over changes in context, perceptual experience is said to exhibit *constancy*, which in turn is taken as supportive evidence that vision is *attempting to compute* a distal scene property (here, lightness). Constancy phenomena therefore provide one of the primary foundations upon which computational level theories of mid-level vision are built. The problem with this justification is that perceptual constancies are always a matter of degree. Perceived lightness, for example, *does* vary across context, surface type, and illumination conditions. How constant must our experience of a distal quantity be to sustain claims that it is a computational goal of the visual system? What kinds of deviations from veridicality provide information that an experience or perceptual ability is epiphenomenal versus a true computational goal? No answers, or method of deriving such answers, currently exist. But it is precisely such answers that are needed to scientifically evaluate whether any proposed natural task qualifies as computational goal of the visual system.

4. Veridicality and the problem of origins

Both Bayesian models and Marr's computational level theory rely on facts about the world to constrain the interpretation of images. Marr referred to these as "physical constraints," whereas Bayesian models require that all possible scene interpretations appear in both the prior and the likelihood; any state receiving zero weight can never be experienced, as it is not one of the candidate explanations of the image data. Despite the central role played by such inbuilt knowledge in explaining perception, essentially nothing is said about how information about the world actually makes into the heads of

animals. In the context of evolution, the typical answer involves little more than a list of ingredients, consisting of "a very long time," "randomness," and "natural selection." Natural selection is charged with the job of implanting knowledge into the heads of animals about the physical states of the world, as well as some understanding of how scene variables interact to create images. This knowledge will putatively favor animals whose heads contain a closer approximation to the true (or useful) states of the world than their less endowed competitors. Perception is thereby rendered to be a rational inference process that selects the best guess about world states on the basis of knowledge acquired over its evolutionary history, although all the explanatory work is performed by something—the priors and likelihoods—whose genesis and existence receive no rational explanation.

What is needed to provide a principled account of the genesis of the information needed to interpret "ambiguous" images? The answer to this question depends on how the term "ambiguity" is construed. One view assumes that images are inherently incapable of specifying their source; the only way to resolve this ambiguity is to have a list of the possible true states of the world (the priors) and how those world states map onto images (the likelihood) (see, e.g., Geisler & Kersten, 2002). Note, however, that this view *presumes the very information that is putatively unavailable from the input*; the information needed to *decode* the images is something that could never be *encoded* from the source of the images. It is, therefore, unclear how this information could ever arise from either selection processes or learning.

One way to avoid such problems is to *require* that any knowledge accrued over the course of evolution (or learning) that is used to resolve the "ambiguity" of images *be derivable from properties of the input*. Note that this involves a much weaker sense of "ambiguity" than that discussed above. It may be the case that any *particular* image at a *particular* moment in time is ambiguous with regard to its "source," but when considered over the history of the species and the organism, for the particular environment that a species occupies, the presumption is that there must be sufficient information available to explain *what it is that a species actually derives from the input*. It is this latter point that distinguishes this view of ambiguity from the Bayesian "recovery" or "estimation" view of ambiguity described above. There is no claim that the visual system derives a veridical representation of the world, or even attempts to derive a veridical representation. All that is required is that an animal recovers sufficiently useful information in a manner that increases its fecundity. More specifically, the argument is that there is sufficient information, over phylogenetic and ontogenetic time, to explain *what it is that an animal actually perceives about its environment*, whether or not this maps onto some measure of "ground truth" (see, e.g., Anderson, in press; Marlow & Anderson, 2013; Marlow, Kim, & Anderson, 2012). From this perspective, the goal of perceptual theory is not to explain how the visual system estimates a particular world property, but rather, to explain how the visual system extracts what it *does* about the world on the basis of the input (where "the input" is construed historically).[6]

It may help to distinguish these views by considering a concrete example. Consider the problem of contour interpolation, something that has been extensively studied from a

variety of theoretical perspectives (see, e.g., Anderson, 2007a,b; Anderson & Barth, 1999; Grossberg & Mingolla, 1985). The importance of contour interpolation is typically justified by the ubiquity of camouflage and occlusion in natural scenes, which cause the images of coherent objects to be fragmented in the images. To recover scene geometry in such contexts, the visual system must somehow link disjoint contour segments into coherent objects. In one series of studies, Geisler, Perry, Super, and Gallogly (2001) generated two models of contour grouping based on two kinds of edge co-occurrence statistics. In one model, they measured the *absolute* edge co-occurrence statistics, which capture the angular and distance relationships between pairs of edge elements in the images. In the Bayesian model, they constructed a grouping function based on the edge statistics of contour elements that belong to the same contour in the world. The key difference is that the absolute edge statistics do not contain any explicit information about whether contours belong to a common environmental source, whereas the Bayesian statistics do. Geisler et al. conducted psychophysical experiments to test whether these derived grouping functions could predict the detection of contours in noise patterns. Surprisingly, Geisler et al. found that absolute edge co-occurrence statistics were equally good at predicting their psychophysical grouping data as the Bayesian model. In other words, a model based solely on the relative occurrences of pairs of edges in the images predicted how contour interpolations occurred in their psychophysical experiments. The potential implications of this finding are potentially far reaching: If the grouping function the visual system uses to link contours can be derived from processes that work directly on properties of the input images, there is no need to invoke the conceptual machinery embodied in Bayesian models: no internalized model of the true states of the world, and no complex model of how images are generated on the basis of those states.

5. Reductionism and explanation

One recurrent theme of the contributions in the present issue involved Marr's concern that reductionist accounts focus solely on understanding *what* an information-processing system does and *how* it does it, and they fail to explain *why* a particular process exists or computes what it does. The arguments I developed above were designed to understand the evolutionary context that lends scientific legitimacy to "why" questions in biology and psychology, and the pragmatic difficulties that are encountered in attempting to articulate and evaluate hypotheses at this level of analysis (see also Bechtel, this volume, on the issue of why questions). The majority of papers in this special issue acknowledge the necessity, or at least the utility, of computational level analyses. One notable exception was Bickle, who asserts that Marr's arguments against reductionism only apply to the particular kind of neuroscientific explanations that were constructed during his lifetime. Bickle argues that modern neuroscience offers true "explanations" of system behavior because they allow for a causal analysis that links brain to cognitive function. Although it is clearly true that neuroscience methods have progressed tremendously since Marr's death, they do not provide any greater understanding of the "why" questions that

computational level theory was developed to provide. Consider, for example, the simple phototropic organism that we considered above. A (complete) causal explanation of this organism's behavior would proceed by specifying how light is transduced, and the chemical and/or neural cascades that ultimately result in the species moving in a particular direction (say, away from the light). Such analyses would allow us to predict the animal's behavior, and in that sense "explain" its response, which seems to be the sense that Bickle uses the term "explanation." But does this account explain *why* the animal behaves the way it does? Although a causal analysis allows us to predict this organism's response to light, it provides no understanding of *why this particular behavior* is the one that this species exhibits. Why does this organism move, for example, against the gradients of light into less illuminated regions? The answer to this question lies not in providing a detailed analysis of the causal sequence that generates this behavior; rather, it requires understanding *why such behavior is adaptive*, that is, what *function* the behavior serves in the context of an animal's niche. Functional analyses require more than simply constructing a casual chain that describes what happens and how it happens; they require knowing something about the environment in which a species evolved and the selection pressures that led that particular response to being selected (say, a photophobic response that arose from the correlation between light intensity and predator density). In Marr's language, it requires understanding the "problem" for which this behavior counts as a solution (here, avoiding predators). More generally, it requires understanding the fitness function that was responsible for shaping the particular computational competencies and/or behaviors exhibited by a species. Answering such questions is not simply a matter of developing more sophisticated techniques for establishing causal sequences of brain behavior, no matter how many pages of *Nature* or *Science* are dedicated to such advances.

6. Conclusions

The main charge of Marr's computational level theory was that a truly explanatory understanding of vision—one that answered a question of *why* vision does what it does, rather than merely *what* it does—should begin by specifying its purpose or computational goals. The preceding attempted to clarify what this involves theoretically, and what is entailed in any attempt to pursue this as a research strategy. From a theoretical point of view, computational level theory presumes the ability to determine and evaluate the fitness function that shaped ("engineered") our visual systems to perform particular functions. Given both the richness of the structure and dimensionality of visual experience and our visual abilities, it is far from evident how to determine which aspects of vision were direct products of natural selection, which is required to justify the teleological structure embodied in computational level theory. However, this does not mean that computational level theory should be abandoned; any attempt to do so leaves "why" questions unanswerable. The main issues confronting the invocation of computational level theory are pragmatic: There must be some principled means of formulating computational level hypotheses, and some method for evaluating whether a hypothesized computation level

theory was the product of natural selection. Although Bayesian ideal observer theory provides some means of establishing an upper bound on particular computational level problems, there is currently no criteria by which the *discrepancies* between ideal model behavior and human behavior can be utilized to decide whether a hypothesized "natural task" or "computational goal" was something that was shaped by natural selection, and hence, actually qualifies as a "computational goal" of the visual system. Finally, I argued that any attempt to explain perception as a form of perceptual inference must articulate how the information used to disambiguate images arose over phylogenetic and/or ontogenetic time. Such considerations suggest that the information used to disambiguate images must be derivable from the sensory input over the history of the species and/or the animal. This suggests an alternative account of the "problem" of perception, shifting the focus from understanding perception as an estimation or recovery process of objective world properties, to a focus on understanding what it is the animal knows about the world based on the information available in the input (considered as an historical construct).

Notes

1. References, page numbers, and quotes to Marr (1982) refer to his posthumous book *Vision*.
2. There is a rich philosophical literature on the relationship between teleological explanations and processes whose cause or existence is owing, at least in part, from their consequences (see, e.g., Wright, 1976; Millikan, 1999). It should also be noted that, so expressed, the extension of any such "consequence etiology" is readily extended to the domain of learning.
3. Perhaps one of the best examples is Marr's computational level description of ganglion cells as differential operators. Marr found it impossible to understand the role of "color opponent" ganglion cells from this perspective, and on the basis of his computational level theory even declared "no receptive fields should mix color and spatial variations in the manner of Figure 3-85(a)" (p. 264). The figure in question depicts a R+G- center/surround ganglion cell architecture, which despite Marr's protestations, are experimentally well established.
4. An entirely separate paper could be written on the conceptual issues raised by this view. At the least, the presumption that evolution provides a list about the states of the world that it uses to interpret the images begs the very question perceptual theory is intended to answer, an issue I return to below.
5. This can be construed as a form of scientific "Jeopardy," after the TV game show that requires contestants to derive questions from statements, such that a given statement serves as the answer to the inferred question.
6. This view shares some features of J. J. Gibson's ecological approach to vision. The main difference is that Gibson argued that there was sufficient information in the optic array to uniquely specify particular properties of the world. The view offered here offers a more tenuous link with the world; the claim is that there is sufficient

information in the input, considered historically, to specify what it is that an animal *knows* about its environment, even if that knowledge in some sense non-veridical.

References

Anderson, B. L. (2007a). The demise of the identity hypothesis, and the insufficiency and non-necessity of contour relatability in predicting object interpolation: Comment on Kellman, Garrigan, and Shipley (2005). *Psychological Review*, *144*(2), 470–487.

Anderson, B. L. (2007b). Filling-in theories of completion: Rejoinder to Kellman, Garrigan, Shipley, and Keane (2007) and Albert (2007). *Psychological Review*, *114*(2), 509–527.

Anderson, B. L. (in press). The perceptual representation of transparency, lightness, and gloss. In J. Wagemans (Ed.), *Handbook of perceptual organization*. Oxford, UK: Oxford University Press.

Anderson, B. L., & Barth, H. C. (1999). Motion based mechanisms of illusory contour synthesis. *Neuron*, *24*, 433–441.

Anderson, B. L., O'Vari, J., & Barth, H. (2011). Non-Bayesian contour synthesis. *Current Biology*, *21*, 492–496.

Bechtel, W., & Shagrir, O. (2015). The non-redundant contributions of Marr's three levels of analysis for explaining information processing mechanisms. *Topics in Cognitive Science*, *7*(2), 312–322.

Bickle, J. (2015). Marr and reductionism. *Topics in Cognitive Science*, *7*(2), 299–311.

Geisler, W. S., & Diehl, R. L. (2003). A Bayesian approach to the evolution of perceptual and cognitive systems. *Cognitive Science*, *27*, 379–402.

Geisler, W., & Kersten, D. (2002). Illusions, Perception, and Bayes. *Nature Neuroscience*, *5*, 508–510.

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*, 711–724.

Gould, S. J., & Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proc. Roy. Soc. London B*, *205*, 581–598.

Grossberg, S., & Mingolla, E. (1985). Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentations. *Perception and Psychophysics*, *38*(2), 141–171.

Lewontin, R. C. (1996). Evolution as engineering. In J. Collado-Vides, B. Magasanik, & T. F. Smith (Eds.), *Integrative approaches to molecular biology*. Cambridge, MA: MIT Press.

Marlow, P. J., & Anderson, B. L. (2013) Generative constraints on image cues for perceived gloss. *Journal of Vision*, *13* (14:2), 1–23, doi: 10.1167/13.14.2

Marlow, P. J., Kim, J., & Anderson, B. L. (2012). The perception and misperception of specular reflectance. *Current Biology*, *22*, 1909–1913.

Marr, D. (1982). *Vision*. San Francisco: W.H. Freeman and Co.

Millikan, R. G. (1999) Wings, spoons, pills and quills: A pluralist theory of functions. *Journal of Philosophy* *96*.4: 191–206.

Rosen, R. (1985). *Anticipatory systems: Philosophical, mathematical, and methodological foundations*. New York: Pergamon Press.

Wright, L. (1976) *Teleological explanations*. Berkeley and Los Angeles: University of California Press.