Marr's Theory of Vision and the Argument From Success[1]

Peter A. Morton

University of Western Ontario

1. Introduction

A central aspect of the computational theory of vision developed by Marr and his coworkers is the use made of contingent regularities in the physical environment to explain how the visual system determines the shape and location of objects in the world on the basis of the spatial organization of the retinal image. Marr (1982) refers to these environmental regularities as "natural constraints" and "physical assumptions."[2] In this paper I am concerned with recent arguments concerning the implications of this feature of Marr's theory for understanding psychological explanation. Burge (1986) claims that Marr's use of natural constraints shows the theory to be *nonindividualistic*, in the sense that the individuation of psychological states depends essentially on the objects and conditions of the world external to the subject. While it does capture an important aspect of Marr's theory, Burge's account is misleadingly incomplete. To see this it is useful to compare the issue of individualism with the question whether the theory is compatible with *methodological solipsism* as a research strategy. Methodological solipsism is the rule that "no psychological state, properly so-called, [should presuppose] the existence of any individual other than the subject to whom that state is ascribed" (Putnam 1975, p. 220). Fodor (1980) maintains that methodological solipsism, realized as the assumption that *semantic* properties of psychological states are irrelevant in explanations of behaviour, defines the approach to theory construction of computational psychology. I will argue that while in some sense Marr's theory is nonindividualistic in the way Burge suggests, it does not follow that it represents an exception to methodological solipsism as a research strategy in cognitive science. The argument extends to Kitcher's (1988) contention, argued on grounds similar to those advanced by Burge vis a vis individualism, that Marr's theory violates methodological solipsism. But the disinction between methodological solipsism and individualism is important for recognizing a crucial aspect of Marr's work. As we shall see in more detail, Burge's view is similar in motivation to Gibson's (1966) theory of direct perception. On this theory, the visual system "picks up" information in the ambient light. My contention will be that methodological solipsism is satisfied by Marr's theory for precisely the reasons that Marr has objected to Gibson, so that we will understand the theory completely only if we see what Burge's argument fails to take into account.

PSA 1988, Volume 1, pp. 154-161

https://doi.org/10.1086/psaprocbienmeetp.1988.1.192981 Published online by Cambridge University Press

2. Physical Assumptions and Natural Constraints

Let us begin with a brief summary of the problem Marr intends natural constraints and physical assumptions to solve in theories of vision.

Marr's goal is a theory that will explain how the human visual system determines the presence and position of objects in the world on the basis of information available in the retinal image. Marr argues that if a theory is to explain this process it must state what information in the image is utilized in perception. It must also give a physiologically realistic description of how that information is extracted and manipulated. The computational approach to vision accomplishes these goals by describing the process as a series of transformations that begin with light intensity values of the image, and terminate with an appropriate description of the environment as output.

Marr contends that a complete computational theory has three explanatory levels: computational, algorithmic, and physical levels of explanation. The *computational level* describes vision as an information processing task. The goal at this stage is to determine which properties of the image are used to describe the properties of the physical environment. The computational level of a theory shows how a description of the distal stimulus can be obtained as a function of specific features of the image. The input, output, and intermediate stages of the visual process are claimed to be symbolic codes in some sense, so that the primitive symbols of the system encode properties of the image that are used to determine properties of the physical environment. The computational level involves the specification of rules defined over the symbolic codes that the transformations must satisfy to culminate in a correct description. The goal of the *algorithmic level* of theorizing is the design of algorithms that implement the rules postulated at the higher level. Marr repeatedly emphasizes that there will be a number of algorithms that satisfy the rules comprising a given computational description of a visual task. (See for example Marr 1982, p. 116f.) The *physical level* of theorizing specifies how these algorithms are realized in neurological structure.

The isolation of *natural constraints* serves the first, computational, level of theory construction. Marr uses the term to refer to *global* properties of the physical environment that can be used to eliminate potential ambiguities in interpreting a representation. Marr's preferred example is the set of constraints utilized in his explanation of stereopsis. The visual system determines the distance of physical objects from the eye by measuring the differences between the positions of the images of an object on each retina. In order for this calculation to be performed it is necessary to determine which retinal features correspond to the same object. Marr contends that this problem can be solved if the operation of the visual system makes two implicit assumptions about the physical world: (1) a given point on a physical surface can occupy only one position in space at any one time; (2) matter is cohesive and separated into objects with generally smooth surfaces. These two natural constraints are translated into constraints on the allowable correlations of symbol tokens in the representations of the image obtained from each eye:

(1) *Uniqueness*. Each descriptive element (i.e., each symbol token) can match at most one item in the representation of the other image.

(2) *Continuity*. The disparity of the matches varies smoothly almost everywhere.[3]

Marr's computational theory of stereopsis is essentially a proof that if a matching of features in the two images satisfies the constraints on allowable matches (and if the images are sufficiently detailed), the correspondence will be correct.

The role of *physical assumptions* is to aid in the specification of the primitive symbol tokens utilized by the system. Like natural constraints, physical assumptions are assumptions concerning global properties of the physical world. But their role is more general than that of natural constraints. Rather than supporting a particular operation like stereopsis, physical assumptions are used to isolate those features of the retinal image that are reliable indicators of surface geometry. The demonstration of a need for an explicit isolation of such features is a characteristic element of Marr's approach to vision. Physical assumptions are designed to ensure that the properties of the image encoded by the primitive tokens of the system are sufficient for detection of the shapes and surfaces in the environment. Spatial features of the image that Marr suggests are so encoded include: (1) average local light intensity; (2) size, density, and orientation of items in the image that are themselves similar to one another in size, local contrast, colour, and spatial organization; (3) distances and orientation of neighbouring pairs of similar items. At the computational level a theory must demonstrate that the features encoded by the primitive symbols are sufficient to determine surface geometry from the image. Physical assumptions are used to establish this conclusion.

3. Individualism and the Theory of Vision

With this outline of Marr's use of natural constraints and physical assumptions in mind, let us consider the claim that the theory is nonindividualistic. Burge's argument for this claim rests on two facts. (1) Marr's theory is designed to explain our *success* at certain visual tasks, so that the computational states are specified in a way that will account for the *veridicality* of our visual representations. (2) The explanations the theory provides assume that the visual system has evolved so as to utilize contingent facts about the environment; our abilities are thus taken to require that natural constraints ensure the validity of such "visual rules" as Uniqueness and Continuity in stereopsis. It follows that the shape of the theory -- including the character of the representational states involved -- depends to a large extent on facts about the external world. Let us look first at what is right in Burge's description of Marr's theory.

On Marr's view, the explanation we give of visual perception depends crucially on the nature of the environment. This is an important consequence of Marr's argument for a computational level of description. The problem at the computational level is to understand how veridical representations of the world can be obtained from light intensity values in the image. The solution to this problem is to find features of the image, light intensity gradients or patterns in the geometrical structure at different scales for example, that carry reliable information about the world, and from which the final representation can be built. One of Marr's insights into the problem is that only a study of the contingent regularities in the actual environment will reveal which image properties are reliable carriers of useful information. Notice that Marr's emphasis on the importance of concentrating on the structure of the real world shares much with Gibson's (1966) search for higher-level invariants in the ambient light. On Gibson's view the visual process consists of extracting information from the surrounding light by detecting invariants like texture gradients in the flow of proximal stimulation. Marr considers this approach the nearest anyone has come to the computational level of theory. He says,

> Gibson's important contribution was to take the debate away from the philosophical considerations of sense-data and the affective qualities of sensation and to note instead that the important thing about the senses is that they are channels for perception of the real world outside or, in the case of vision, of the visible surfaces. He therefore asks the critically important question, How does one obtain constant perceptions in everyday life on the basis of continually changing sensations. This is exactly the right question, showing that Gibson correctly regarded the problem of perception as that of

> recovering from sensory information "valid" properties of the external world. (Marr 1982, p. 29)

So Burge is right in pointing out that we cannot construct a theory of vision like Marr's if we restrict our attention to the subject and his sensations. Understanding perception requires a knowledge of how the patterns of retinal stimulation are causally linked to appropriate features of the world. The content of representations postulated to explain perception depends on the structure of the physical environment, and this is justified by the presumption that we have evolved so as to exploit regularities in our surroundings that connect the structure of light intensities with shapes and surfaces in the world. In order for states of the system to carry information about conditions in the world, it must be the case that they are regularly correlated with such objective conditions. Since the representations postulated by Marr's theory are determined by their causal connections with the world, it follows that in the normal case they will be veridical.

For Burge, all of this suggests a thought experiment similar to Putnam's Twin Earth story. We can imagine there being two individuals who have identical physical histories but who live successfully in environments that differ in the regularities that underlie the individuals' visual abilities. Burge argues that the dependence of Marr's theory on the specification of causal relations between representations and contingent conditions in the world requires the ascription of type-distinct representational states to the two individuals, *despite their physical type-identity*:

> The methods of individuation and explanation are governed by the assumption that the subject has adapted to his or her environment sufficiently to obtain veridical information from it under certain normal conditions. If the properties and relations that normally caused visual impressions were regularly different from what they are, the individual would obtain different information and have visual experiences with different intentional content. (Burge 1986, p. 35)

Burge concludes that on Marr's approach to vision, representational states of a subject do not supervene on physical or functional states. The former will vary as we (counterfactually) alter the description of the surrounding environment while holding the subject's physical history fixed; hence, Marr's theory is nonindividualistic.

The problem with this account is that it overlooks a crucial feature of Marr's theory. Not only is it necessary on Marr's view to specify the informational links that produce veridical perceptions, we must also explain how that information is processed within the system. In computational terms, this requires that we specify the rules according to which representations are constructed. Here it is useful to consider Gibson, since it is precisely this requirement that is overlooked in Gibson's theory, and it is precisely because this requirement is overlooked that Marr rejects Gibson's reduction of theory of perception to theory of stimulus invariants. As Gibson sees it, the task of explaining vision ends with the discovery of stimulus invariants associated with the appropriate properties, and there is no need to determine how the invariant relations are specified within the system. As the title of Mace (1977) nicely describes it, the task is to "ask not what's inside your head, but what your head's inside of." The inadequacy of this approach for Marr and other computationalists, is that it fails to recognize that the detection of invariants is an information processing problem, and a complex one at that. (Marr 1982, p. 30) To point out the common error of failing to recognize this Marr cites Austin's (1962) claim that there are no rules or procedures to determine the real shape of a perceived object, and observes that,

> There are ways of describing the shape of a cat to an arbitrary level of precision, and there are rules and procedures for arriving at such descriptions. This is exactly what vision is about, and precisely what makes it complicated. (Marr 1982, p. 31)

It is this insistence on the specification of rules by which the representations are built up that underlies the theory's commitment to methodological solipsism.[4] To see this, let's look once more at the theory of stereopsis.

The Uniqueness and Continuity constraints in the solution to the matching problem are rules stated over specific representations. The validity of these constraints rests on assumptions about contingent physical regularities. The natural constraints ensure that *normally* the operation of the rules will yield veridical output. Were this not the case the rules would not be deployed. This is just the point of Burge's argument. But the *operation* of the rules is insensitive to the truth or falsity of their output, and this is just the point of methodological solipsism as a research strategy. So while the theory is nonindividualistic in the sense that the content of the rules is determined in part by regularities in the environment, the demand that these regularities be exploited in the theory by rules that have no access to the truth-value of the representations commits the theory to methodological solipsism.

The point becomes even clearer when we look at the role of physical assumptions. Regularities captured by these assumptions ensure that the encoded properties of the retinal image suffice for building up reliable descriptions of the world. In the construction of the theory, physical assumptions determine which aspects of the retinal stimulus are represented by the primitive symbols of the system. While natural constraints ensure the validity of the computational transformations, physical assumptions ensure that the descriptive elements over which the rules are stated preserve the appropriate properties and relations in the stimulus. It is the physical assumptions that guarantee that only these elements, and the application of the rules, need be appealed to in the theoretical explanation of how descriptions of the visible world are arrived at.

There is one more similarity between Burge's argument against individualism and the debate between Gibson and computationalists. Gibson's argument for a need to concentrate on how an organism is successful at perceptual tasks leads him to be critical of the introspectionist emphasis on illusions and nonveridical perceptions. This emphasis he attributes to a narrow and undermotivated notion of the stimulus. It is Gibson's view that if we define the stimulus instead to consist of higher-order relations, and if we suppose it to be structured over time, then cases of perceptual error will be revealed as situations in which some part of the stimulus is absent. This approach is intended to avoid the conclusion that perception involves inference, argued for on the grounds that perceptual error can only be explained as error of inference. The problem with Gibson's account of misperception, as Fodor and Pylyshyn (1981) point out, is that it cannot account for situations in which the same stimulus information results in veridical perception in one subject but not in another. The need to explain perceptual error *as well as* the general reliability of perception also forms the basis for methodological solipsism in perception theory.

Burge correctly points out that Marr's explanation of our success at visual tasks presupposes the existence of causal relations that result in veridical perceptions under normal circumstances. This adds some truth, he says, to the slogan that "mistakes presuppose a background of veridicality." (Burge 1986, p. 41) This may be so, but it is still the case that in order for the theory to be genuinely explanatory it must be possible for representations to be false. Hence, the content of our visual representations (that aspect over which the rules are stated) must remain constant in situations where their normal cause is absent. And this is just to say that the content of representations is insensitive to changes in their correlations with external conditions. So unless methodological solipsism is adhered to in perception theory we will be unable to explain why we sometimes misperceive objective states of affairs.

In a discussion of Marr's theory, Kitcher gives the following argument against the conclusion of this section.

> Most directly, if Marr is right, then a theory of vision must incorporate information about the environment, both in describing the representations produced by the system and in describing the constraints that it uses to disambiguate information in the gray-level array. Thus, Marr's project violates Fodor's canon of Methodological Solipsism, because it does not confine itself to syntactic or formal features of internal representations; rather it makes essential reference to factors beyond the subject's skin in characterizing psychological states. (Kitcher 1988, pp. 13-4)

Kitcher is correct in her claim that Marr's theory is shaped by facts about the environment. But as we have just seen, the explanation the theory gives of the operation of the system appeals only to nonsemantic properties of representations. Kitcher's argument fails to distinguish the use of environmental regularities to explain the success of visual mechanisms from the description of those mechanisms in the theory itself. Marr's commitment to methodological solipsism is manifest in his insistence on the latter.

4. Fodor on the Argument from Success

Fodor (see his 1987, pp. 33-44) has given a similar response to Burge, although he puts the point in somewhat different terms. The issue of individualism as it is commonly discussed is the general question whether relational properties of mental states are relevant to psychological taxonomies. This question has been confused with the issue of methodological solipsism as a research strategy. The real point in the defence of individualism, according to Fodor, is not the relevance of relational properties; rather it is the contention that psychological taxonomies should not include distinctions that are not *causally* relevant. This requirement is constitutive of scientific taxonomies generally. A taxonomy is individualistic as Fodor uses the term if it "distinguishes between things insofar as they have different causal properties, and . . . groups things together insofar as they have the same causal properties." (Fodor 1987, p. 34) A causal property is one that an object has in virtue of which it is subsumed under a causal generalization. There are individualistic taxonomies in this sense that include relational properties -- *being a planet*, for example, is a relational property that can appear in causal explanations. Methodological solipsism is just the principle that the individuation of mental states by semantic properties is not among such taxonomies.

But when applied to vision theory, this account of the individuation of states does not give a full explanation of the role of the appeal to success in visual tasks emphasized by Burge. The lesson learned from Gibson is that the way to explain visual ability is to look at the relations between the organism and the environment. As Burge suggests, these relations are utilized in Marr's work through appeals to the *truth-preserving* property of certain computational transformations given the character of the natural environment. Fodor recognizes this aspect of theories like Marr's; however Fodor takes it not to be a violation of his version of individualism, but as a demonstration that such explanations represent an exception to methodological solipsism.

These sorts of explanations square with *individualism*, because the relational facts they advert to affect the causal powers of mental states; indeed, they affect their very existence. But naturally, explanations of this sort -- for that matter, *all* teleological explanations -- are ipso facto nonsolipsistic. (Fodor 1987, p. 44)

The point however cannot be that the theory of vision is both individualistic in Fodor's sense and nonsolipsistic. If we take this approach, Marr's theory will yield two taxonomies of representational states depending on our explanatory goals. If we take causal properties of these states to be specified by their role in the production of behaviour, then since the behaviour of the visual system is unaffected by the semantic values of its representations, semantic properties will not be included. Thus representations will not be individuated by semantic properties if the taxonomy is designed solely to account for the *behaviour* of the system. But as Burge has shown, if our purpose is to explain why this behaviour is *successful* we *will* have to specify the content of representations to account for their veridicality. Thus we will pull the theory two ways depending on whether we count the reliability of perception to be part of what it is designed to explain.

However this response is unnecessary. The apparent problem is that while we want to allow for the theoretical utility of semantic properties, we do not want to go so far as to admit truth values into explanations of behaviour. But this is not a problem at all: Methodological solipsism is the requirement that truth values must not enter explanations of behaviour. But this requirement is unaffected by what for Burge is a nonindividualistic taxonomy of states. We can allow that representational states do not supervene on physical states without thereby sacrificing the explanatory power of computational explanations by requiring that the *operation* of the system be indifferent to semantic values. Fodor is quite right in conceding that "teleological" arguments that justify ascription of perceptual mechanisms are nonsolipsistic. But the claim that these arguments affect the design of the system, including the individuation of states, does not preclude the possibility that the theoretical explanation of the operation of the visual system is entirely solipsistic. The point of Section 3 above is that a complete understanding of Marr's work must take into account both aspects of the theory.

Notes

1I would like to thank William Demopoulos for many helpful suggestions and improvements. G. Keith Humphrey first directed my attention to Gibson.

2In this paper I will distinguish between natural constraints and physical assumptions. Marr's usage of the two terms varies but it is useful to keep two distinct notions as we shall see.

3To be more exact, there are three matching constraints. The first, which I have not mentioned is that two descriptive elements can match only if they could have arisen from the same physical markings. This requirement, which Marr calls *compatibility*. does not depend on either of the two natural constraints but rather (as Burge notes) on an unremarked assumption that both eyes will produce similar descriptions of the same physical feature.

4The formulation of rules by which information is processed is a concern at the computational level of theory construction, and is not merely a matter to be dealt with at the algorithmic level. Marr makes this clear in his discussion of Chomsky's transformational grammar. He points out that although Chomsky's transformational

https://doi.org/10.1086/psaprocbienmeetp.1988.1.192981 Published online by Cambridge University Press

rules are sometimes criticized because they cannot be implemented on a computer, the theory is in fact "concerned solely with specifying what the decomposition of an English sentence should be, and not at all with how that decomposition should be acheived." (Marr 1982, p. 29) Kitcher's criticism of Fodor's account of computational theories (Kitcher 1988, pp. 4) appears to overlook Marr's insistence on rules at the computational level.

References

Austin, J.L. (1962). *Sense and Sensibility*. Oxford: Clarendon Press.

Burge, T. (1986). "Individualism and Psychology." *The Philosophical Review* 95: 3-45.

Fodor, J.A. (1980). "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology." *Behavioural and Brain Sciences* 3. Reprinted in *Representations: Philosophical Essays on the Foundations of Cognitive Science* pp.225-53. Cambridge: M.I.T. Press.

______. (1987). *Psychosemantics*. Cambridge: M.I.T. Press.

Fodor, J.A. and Pylyshyn, Z.W. (1981). "How Direct is Visual Perception? Some Reflections on Gibson's 'Ecological Approach'." *Cognition* 9: 139-96.

Gibson, J.J. (1966). *The Senses Considered as Perceptual Systems*. Boston: Houghton Mifflin.

Kitcher, P. (1988). "Marr's Computational Theory of Vision." *Philosophy of Science* 55: 1-24.

Mace, W.M. (1977). "James Gibson's Strategy for Perceiving: Ask Not What's Inside Your Head But What Your Head's Inside Of." In *Perceiving, Acting and Knowing*. Edited by R. Shaw and J. Bransford. New Jersey: Lawrence Erlbaum.

Marr, D. (1982). *Vision*, New York: W.H. Freeman.

Putnam, H. (1975). "The Meaning of 'Meaning'." In *Mind, Language and Reality: Philosophical Papers* Volume II. Cambridge: Cambridge University Press.