Supervenience and Computational Explanation in Vision Theory
Author(s): Peter Morton
Source: *Philosophy of Science*, Vol. 60, No. 1 (Mar., 1993), pp. 86-99
Published by: The University of Chicago Press on behalf of the Philosophy of Science Association
Stable URL: http://www.jstor.org/stable/188456
Accessed: 30-07-2017 03:45 UTC

SUPERVENIENCE AND COMPUTATIONAL EXPLANATION IN VISION THEORY*

PETER MORTON†‡

Department of Humanities
Mount Royal College

According to Marr's theory of vision, computational processes of early vision rely for their success on certain "natural constraints" in the physical environment. I examine the implications of this feature of Marr's theory for the question whether psychological states supervene on neural states. It is reasonable to hold that Marr's theory is nonindividualistic in that, given the role of natural constraints, distinct computational theories of the same neural processes may be justified in different environments. But to avoid trivializing computational explanations, theories must respect methodological solipsism in the sense that within a theory there cannot be differences in content without a corresponding difference in neural states.

1. Introduction. An important feature of David Marr's theory of vision is the assumption that early visual processes make use of certain global properties of the physical environment, which Marr calls "natural constraints". This assumption makes it possible to explain how highly rigid, modular processes might obtain detailed information about the stimulus. Recent articles (Burge 1986, Fodor 1987, Kitcher 1988, Egan 1991) have drawn conflicting implications from this aspect of Marr's work for the question whether psychological states supervene on neural states. Yet none, I think, have formed a proper idea of how the issue should be understood in the context of Marr's theory independently of more general questions in philosophy of psychology.

In broad terms, the rule that psychological states *supervene* on neural states is the requirement that any distinctions drawn between psychological states correspond to some difference in neural states. Marr's theory is a *computational* explanation of a specific set of perceptual abilities; the success of neural activity in obtaining information from the retinal image is explained by describing the visual system as a computing mechanism. Consequently, certain states of the system are interpreted by the theory as representing features of the stimulus. The question of supervenience

*Received October 1991; revised March 1992.

†I would like to thank G. Keith Humphrey, David Martens, and especially William Demopoulos for their helpful comments.

‡Send reprint requests to the author, Department of Humanities, Mount Royal College, Calgary, Alberta T3E 6K6, Canada.

Philosophy of Science, 60 (1993) pp. 86–99

in this context is whether states of the system partitioned by their representational content under a computational description supervene on states partitioned by their neurological descriptions.

It is, of course, hard to give a broad account of the representational content of psychological states. However, in the restricted framework of computational theories of low-level processes we can restrict ourselves to a narrow characterization. In standard computation theory, we give a mechanism a computational description by specifying an interpretation function that assigns physical states of the mechanism to elements in the domain and range of a function F, the function computed by the system. Let C be the set of physical states of a mechanism M in the domain of an interpretation function I. We say that *M computes F* if there is a physical-state transition map, $P:C \rightarrow C$ such that, for $c \in C$,

$$I(P(c)) = F(I(c)).$$

If we let the representational content of a state be its interpretation under a computational description, then supervenience is a constraint on interpretation functions to the effect that the same neural state cannot be mapped to more than one representational state.

By the criterion just stated, the claim that a mechanism computes a particular function is a very weak one. With no restrictions on interpretation functions and physical-state transition functions, every physical system computes every function (see Stabler 1987 for a discussion of this point). An important question, then, is the nature of the constraints on interpretation functions that render computational descriptions *explanatory*. The issue of supervenience as it applies to Marr's theory of vision forms a part of this question.

Burge (1986) argues that Marr's use of natural constraints shows the theory to be *nonindividualistic*. By this he means that the content of representational states within the theory depends essentially on objects and conditions of the world external to the subject. According to Burge, the assignment of representational content in Marr's theory will vary with changes in the description of the surrounding environment while the subject's physical description remains fixed. I will argue that in a narrow and specific sense Burge is correct: In some contexts Marr's explanatory framework may justify assignment of different interpretations to systems that have identical physical descriptions but operate in different environments.

However, at the risk of trivializing computational explanations, it is important to recognize that there are restrictions on the ways in which supervenience can be relaxed, and in this respect Burge's account is misleading. I will argue that Marr's theory is committed to *methodological solipsism* as formulated by Fodor, and that this fact provides important constraints on relaxation of supervenience. Briefly, while Burge's argument shows that distinct computational theories with different interpretation functions may be justified for the same physical system in different environments, methodological solipsism requires that *within a theory* identical physical states must be given identical interpretations. Unfortunately, Fodor's description of the role of methodological solipsism in theory construction suffers from a confusion of issues, which has generated unnecessary resistance to his thesis.

2. The Argument from Success. Let us consider Burge's argument that Marr's theory is nonindividualistic. The argument rests on two broad claims about Marr's theory (Burge 1986, 29–33).

(*A*) The theory assumes that the visual system has evolved so as to exploit global regularities, or natural constraints, in the normal environment. The computational processes of vision are truth preserving only if these natural constraints hold, and so our visual abilities are taken to depend in part on the character of the physical world. In this way, relations between states of the world and representational states of the system depend on certain contingent facts about the normal environment.

(*B*) The purpose of the theory is to explain our *success* at certain visual tasks. To this end, the content (and hence the individuation) of representational states is determined by their normal causal antecedents. Thus, the theory specifies representational states in such a way that the contents of visual representations are true in the normal case.

Burge argues from these two premises to nonindividualism of the theory by constructing a thought experiment similar to Putnam's (1975) Twin-Earth story. Given (*A*), it is possible for there to be two individuals who are physically identical but who live successfully in environments that differ with respect to the natural constraints underlying their respective visual abilities. The point is that unless the representational states of the subjects are assigned different contents, one of them would have representations that are regularly false, contrary to (*B*). Burge concludes that Marr's theory requires the ascription of different representational states to the two subjects, despite their physical similarity:

> 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. (1986, 35)

Hence, it is argued, on Marr's approach to vision, representational states of a subject do not supervene on neural states.

Properly understood, both (*A*) and (*B*) are true. The problem that Marr's theory addresses is how veridical representations of the physical world are obtained from light-intensity values in the image. Thus, the representations postulated by the theory are those required to explain the success of visual processes. According to Marr, the early visual system computes a sequence of functions, each of which corresponds to a relation between specific magnitudes in the stimulus. Beginning with light intensities in the image, earlier computations in the sequence provide the input to later computations until the sequence culminates in a single final representation of the layout of objects in the distal scene. The computations that comprise this sequence are highly modular, operating independently of higher cognition and with very limited information about the current stimulus. On Marr's theory, the system is successful in extracting information in this way by exploiting natural constraints in the physical environment.

The initial inspiration in this regard is the work of Gibson. Gibson noticed that while a static retinal image does not suffice to determine the distal scene, significant information about the world can be derived from the *rates of change* of certain stimulus magnitudes with change in the position of the viewer. For example, the three-dimensional shape of an object can be determined from the variation in its two-dimensional retinal projection as the object moves with respect to the viewer. Importantly, the information carried by rates of change of stimulus magnitudes is a consequence of certain uniformities in the structure of the environment. Higher-order variables can be exploited in the manner Gibson suggests because there are regular spatial relations between identifiable features of surrounding surfaces that result from the relatively uniform character of the physical world. For example, as Ullman's (1979) shape-from-motion theorem makes explicit, the recovery of shape from variation in retinal projection depends on the rigidity of physical objects. Thus an essential part of Marr's general theory is that early vision succeeds in extracting information by exploiting relations between higher-order stimulus magnitudes that result from the structure of the normal environment. The relations between stimulus variables computed by the system, and hence the veridicality of the representations generated, depend on the existence of certain natural constraints.

Egan (1991) criticizes Burge's account of Marr's theory. One of the premises in Burge's argument is that Marr's theory is *intentional*, in the sense that states of the system are taken to represent the world outside the subject. Egan takes issue with this premise on two grounds: (1) A large part of the representations specified by the theory represents features of the retinal image that, as Marr emphasizes, do not denote any particular features of the physical world; and (2) even if the computational sequences were interpreted in terms of features of the distal stimulus, this aspect of the theory is merely adventitious, and not properly part of the theory at all. Let us look at these objections in turn.

Egan is correct that much of the computational work described by Marr's theory succeeds in extracting only aspects of the geometrical structure of the image. But this point is a red herring. Although they represent only the image, and not aspects of the physical world, the processes described are computational and thus have representational content under a computational description. Moreover, natural constraints play an important role even at the earliest stages in ensuring that the features of the image isolated by the system are those that reflect the structure of the physical world rather than results of the imaging process. This is evident in the construction of what Marr terms the "primal sketch", which Egan uses to illustrate her point. For example, formation of the initial primal sketch relies on the assumption that features of surfaces that cause intensity changes in the image are spatially localized. Similar assumptions are exploited in detection of intensity changes by "zero-crossings", and in isolating the geometrical structure of the image by the use of "virtual lines". Thus, an important question remains as to whether the assignment of content to these very early states respects supervenience.

Egan's second claim is mistaken. As she points out, Marr's theory consists of three distinct levels of explanation: computational, algorithmic, and physical. The computational level describes the functions computed by the system, while the algorithmic level describes the particular algorithms employed in the computation, and the physical level specifies the physical-state transitions that implement the algorithms. On her view, the computational-level functions can be described in terms *neutral* with respect to the distal stimulus—for example, as functions over natural numbers. However, it is essential to the role of the computational level that the representational states be interpreted at that level in terms of their causal antecedents in the normal environment. This point warrants close attention.

It is clear from concrete applications, as well as from Marr's expository remarks, that the primary role of the computational level of theory is to demonstrate that the computation of a particular function is sufficient to extract, from information available to the system, all the information required at a particular stage of processing. Such a demonstration must do three things: (1) It must provide a canonical description of the function computed; (2) it must describe the conditions under which such a function yields a true representation of a feature of the stimulus from that of another feature; and (3) it must describe the properties of the physical composition of the environment that make these conditions likely to hold. We can see that each of these is an essential component of the theorems that make up the computational level of theory, and that together they demonstrate the importance of (*A*) and (*B*) in Marr's theoretical framework. In particular, we can see all three components playing a crucial role in the three areas of vision most fully developed by Marr's group: edge detection, stereopsis, and shape-from-motion.

Egan agrees that interpreting states of the system as representing the normal environment is necessary in explaining how the system is successful. But she argues that maintaining the same interpretation across different environments is also essential to explain how the system *fails* outside its usual environment, so that Burge's thought experiment is misguided. Her emphasis on the need to account for perceptual failure is both correct and important. However, we will see that *this* constraint on relaxation of supervenience is met by methodological solipsism, not by individualism.

Egan's claim that it is inappropriate to describe the cognitive system at this stage as intentional in any of the rich senses employed in philosophy of mind is also correct. *Intentionality* is a term more properly applied to belief states where problems like referential opacity arise, and it is implausible that the states postulated by Marr are intentional in this usual sense. Restricting the way in which early vision is taken to involve representational states to the narrower sense taken from theory of computation, outlined in the introduction, does not eliminate the question of supervenience with respect to early vision. Rather, it places the question within a specific context, namely, the nature of the explanatory power of computational theories of cognition.

The question, then, is whether the denial of supervenience is a consequence of (*A*) and (*B*) in the manner suggested by the thought experiment. Against this conclusion, Fodor (1987, chap. 2) maintains that the possibility of mental causation depends on supervenience. According to Fodor, if Burge is right we will be unable to explain how cognitive processes are realized in neural structure. Let us look at Fodor's argument.

3. Assessment of the Argument from Success. Whether psychological theories are individualistic is often discussed as the question whether certain *relational* properties of mental states are relevant to psychological taxonomies: Is the type-identity of the psychological states of a subject affected by relations between the subject and the external world? But Fodor argues that the defense of individualism should not rest on the relevance of relational properties. The question, as he sees it, is whether psychological taxonomies should include distinctions that are not *causally* relevant. According to Fodor, a taxonomy is individualistic if it "distinguishes between things insofar as they have different causal properties, and . . . groups things together insofar as they have the same causal properties" (1987, 34) where a causal property is one in virtue of which an object is subsumed under a causal generalization. He argues—correctly, it would seem—that it is constitutive of scientific theories generally that taxonomies are individualistic in *this* sense. In his view, it follows from this that the nonindividualistic taxonomies suggested by Burge's thought experiments are unscientific.

Fodor's argument runs as follows: Changes in the environmental surroundings of a subject that do not affect the subject's neural states can have no relevance to the causal generalizations into which those states can enter. In particular, distinctions among representational states that do not correspond to differences among neural states cannot affect the *behavior* of the system. Thus, such taxonomies remove the basis for descriptions of mental causation of behavior. So any taxonomy that does not preserve the supervenience of representational states draws distinctions that reflect no differences in the causal properties of representational states. Since this violates the just-noted requirement on scientific theories, supervenience is a first principle of psychological explanation.

But the argument outlined in the previous paragraph is unsound. Fodor is certainly right that the explanations provided by psychological theories cannot appeal to differences in the representational states of a subject that do not correspond to any difference in neural states. Explaining the success of visual processes, for example, is trivialized if we are allowed to alter the content of visual states so as to guarantee their veridicality. This constraint, however, is not inconsistent with nonsupervenience in Burge's thought experiments, as Fodor maintains. The error stems from the fact that Fodor takes the taxonomies imposed by computational theories to be *causal* taxonomies. According to Fodor, representational states are individuated by grouping them according to their causal properties. But the generalizations that underlie *computational* explanations are not causal generalizations in the sense that the taxonomies induced by computational theories rest on sameness of causal powers. The explanatory power of computational theories lies in the equivalence of different physical systems under the same computational description. So individuation in such theories is determined by assignment of content under an interpretation function. According to Fodor, to deny supervenience is to construct a taxonomy that ignores causal properties; but properly understood, the issue is whether there are legitimate explanatory reasons for assigning different computational descriptions to the same physical mechanism.

Trivially, a physical system may fall under more than one computational description since under some description every system computes every function. The thought experiments are intended to show that *good* *explanatory grounds* exist for altering the computational description of a system under different environmental conditions. And we can at least consistently describe such a situation.

Let us suppose that as a consequence of some set of natural constraints in the normal environment, the value of a stimulus magnitude M_1 is a function F of another stimulus magnitude M_0, and that according to a theory Γ, the visual system determines the value of M_1 by computing this relation. If it is a complete theory, Γ describes a physical-state transition function over neural states of the system, and it provides an interpretation function that maps these neural states to values of M_0 and M_1.[1] Burge's thought experiment describes a situation in which a *distinct theory* Γ' for the system describes the same physical-state transition function, but where the interpretation function maps states in the range of the physical-state transition function to values of a magnitude M_1' distinct from M_1. Then Burge's argument is that Γ' will be the preferred theory of the system if the following two conditions are met: (1) The system is in an environment where, given some new set of natural constraints, M_1' is a function F' of M_0; and (2) the system computes F' under an appropriate interpretation function. Since the computation of F' will be part of a sequence of computations that generates a composite representation, we must also suppose that a similar relation holds between M_1' and other magnitudes in the stimulus, and that interpretation functions exist that describe the state transitions of the system as computing these relations.

Concrete situations in which these conditions are satisfied are perhaps difficult to imagine, and Burge does not really provide convincing examples. Yet the conditions do not appear impossible, and so, prima facie, Burge's conclusion is consistent with Marr's explanatory framework. However, the significance of this conclusion lies less in the nonindividualism of the theory itself than in the emphasis it places on Burge's two premises (A) and (B), which *are* important—for there is a danger in identifying nonindividualism as the most significant consequence of (A) and (B) in the way Burge does. Nonindividualism tells us that in certain contexts it is legitimate to assign distinct interpretations to the same neural processes. Burge's account of Marr's theory says that such assignments are warranted in explanations of early vision in order to preserve the overall veridicality of the subject's representations. But clearly we must place constraints on the assignment of interpretations in computational theories to ensure the explanatory power of those theories. The specification of these constraints must be spelled out to determine the nature of compu-

[1]In Marr's theory the values of M_0 and M_1 will be represented by symbolic expressions under an encoding function, and F will be computed by carrying out a symbolic transformation function. But these complications do not alter the point, so they can be ignored here.

tational explanation. For example, as Fodor and Pylyshyn (1981) point out, the trivialization problem that attaches to Gibson's theory of direct perception is that, although he successfully identifies causal links between stimulus features that ensure the veridicality of perception, he does not specify the *computations* by which these causal links are exploited in the mechanical generation of representations. Fodor's concern about the mechanical production of behavior still needs a solution. In the next section, I will claim that methodological solipsism is an important constraint to ensure the explanatory power of computational explanation, and in particular that it sets important limits on the relaxation of supervenience.

4. Supervenience and Methodological Solipsism. It is important to notice that methodological solipsism is presented differently in Fodor's *statement* of the principle than it is in his description of how the principle functions as a constraint on computational explanations of behavior. In his statement of the principle, Fodor presents methodological solipsism as a claim similar to Putnam's original version, and to some extent similar also to Burge's individualism. But the principle as it is described in its application provides the necessary restrictions on supervenience.

In his statement of methodological solipsism, Fodor says that the principle is represented in computational psychology by the "formality condition" which says that "two thoughts can be distinct in content only if they can be identified with relations to formally distinct representations" (1981, 227). By "formal properties" Fodor intends any *nonsemantic* properties of psychological states, where semantic properties include truth, reference, and representational content. So the formality condition says that the type-identity of representational states is not affected by changes in the environment of a system that do not affect the physical or symbolic character of computational states.[2] This much appears as a version of Putnam's original formulation of methodological solipsism, and also of Burge's individualism.

But Fodor's *application* of methodological solipsism concerns different issues than those that motivate either Burge's or Putnam's discussions. Here is Fodor's statement of how methodological solipsism functions in computational psychology:

> I'm saying, in effect, that the formality condition, viewed in this context, is tantamount to a sort of methodological solipsism. If mental processes are formal, then *they have access only to the formal properties* of such representations of the environment as the senses

[2]For Fodor appeal to symbolic transformations is constitutive of computational theories. On his view, then, formal properties are properties of symbol sets. I think, however, that this is not true of very early visual computations such as edge detection.

> provide. Hence, *they have no access to the semantic properties* of such representations, including the property of being true, of having referents, or, indeed, the property of being representations of the environment. (1981, 231; emphasis added)

Notice that in this statement methodological solipsism is an assertion about *the kinds of properties to which mental processes have access*. Fodor's application of methodological solipsism is intended to place constraints on what can enter into descriptions of the domains over which mental processes are defined. The point of methodological solipsism, as Fodor applies it, is that any distinctions between mental states that affect the behavior of the system must correspond to distinctions between physical properties of the system.

By contrast, individualism is not a restriction on the properties of mental states that influence behavior. Rather, it is a constraint on psychological explanation generally, namely, *psychological explanations must not appeal to distinctions among representational states to which the processes of the system have no access*. No conflict arises between methodological solipsism and a denial of individualism, for they are concerned with different things. Methodological solipsism says that the processes of the system have no access to semantic properties of mental states, so that these properties cannot affect the behavior of the system. By denying individualism we allow only that semantic properties of computational states can be appealed to in psychological explanations for some purpose, where how this appeal can appear within the theory is left open. So an appeal to semantic properties that violates individualism is consistent with methodological solipsism so long as it does not appear in the description of the domains over which mental processes are defined.

The appearance of conflict between the two doctrines arises from Fodor's later remarks describing methodological solipsism as restricting the domain of what is relevant to psychological explanations. Thus, for example, Fodor remarks that "there can't be a psychology of knowledge" (1981, 228). The question follows, then, why he broadens the scope of methodological solipsism in this way. I think that the inference to the conclusion that methodological solipsism is a restriction on psychological explanation generally has the following form:

> The business of psychology is to provide explanations of the mental causes of behavior. According to methodological solipsism, the mental processes that generate behavior have no access to semantic properties of representations. Therefore semantic properties are irrelevant to psychology.

However, we have seen that semantic properties *are* relevant to psychological explanation, and this is the thrust of Burge's argument. Indeed, given the role of the computational level of description mentioned in section 2, formulating a "psychology of knowledge" is the first part of theory construction. On the presumption that we have evolved to exploit the effects of natural constraints, the character of cognitive processes is determined in part by semantic properties of representational states. However, this fact does not entail the conclusion that methodological solipsism is false. Rather, it entails that methodological solipsism is a constraint on *how* semantic properties can enter into psychological explanations. In this form methodological solipsism is both true and important.

So we need to specify how semantic properties can enter into computational explanations of perceptual abilities. I think that the answer is now clear. *Semantic properties can appear in explanations of perception as part of the justification of the choice of a computational theory as characterizing a particular mechanism in a particular environment.* This statement explains the basis of nonindividualism. Burge's thought experiment is intended to provide an example of a case in which computational theories of two systems in different environments may differ only in the interpretation function that takes computational states of the system to properties of the environment. *But by methodological solipsism semantic properties cannot appear within a computational theory as part of the individuation of representational states over which the computations are defined.*

Fodor (1987, 43–44) reaches much the same conclusion about relaxation of supervenience. He points out that in justifying the choice of a particular theory, psychologists appeal to features of the world that do not affect the causal properties of psychological states. But, ironically, he draws a conclusion nominally opposed to the one we have just reached. In his view, the facts we have surveyed show that Marr's theory is individualistic but that the justification of computational theory violates 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. (Ibid., 44)

Making the point this way poses two problems. First, Fodor's account rests on his appeal to causal properties, and this is *not* an appropriate way to frame the explanatory form of computational theories. On his view, computational theories are causal theories, so that assignment of representational content cannot violate individualism. As we have seen, on a correct view of computational theories this is not a concern. Second, on

Fodor's version of the issue, theories like Marr's will be ambiguous with respect to the criteria by which representational states are determined. On his view, causal properties can be understood in different ways. If causal properties of psychological states are specified by their role in the production of behavior, then since the behavior of the visual system is unaffected by the semantic values of its representations, semantic properties will not be included. So if the representations are specified solely to account for the *behavior* of the system, they will not be individuated by their semantic properties. A theory of this sort would then appeal only to the formal or syntactic properties of computational states, leaving the theory neutral with respect to the organism's relations with the environment, much as Egan suggests. But if our purpose is to explain why this behavior is *successful* we will have to specify the content of representations in a way that accounts for their veridicality, and this will require specifying causal powers in a way that includes relations with the environment. Thus, the theory will be pulled in two ways, depending on whether we count the reliability of perception as part of what it is designed to explain.

To state things in this way, however, is unnecessary. While we want to allow for the theoretical utility of semantic properties, we do not want to assign semantic properties a role in explanations of *behavior*. Nor is this a problem, for the requirement that semantic properties not enter explanations of behavior is consistent with the requirement that these same properties play a role in the choice of computational theory. We can allow that representational states do not supervene on physical states, in the sense that computational theories of the same physical system may differ only in their assignment of representational content to computational states. We do not thereby sacrifice the explanatory power of computational explanations as long as we require that *within a theory* there are no differences in representational content, and accordingly no differences in truth-values, that do not correspond to a difference between physical states of the system.

The confusion generated by Fodor's way of describing the role of methodological solipsism is revealed in Kitcher's (1988) discussion of Marr's theory:

> 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 grey-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. (Pp. 13–14)

Kitcher's argument is clearly directed at Fodor's extension of methodological solipsism to a denial of the relevance of semantic properties in psychological theories generally. Notice that, as Kitcher understands it, methodological solipsism is the assertion that psychology must restrict itself to the "syntactic or formal features" of representations. On her reading, methodological solipsism denies the computational level of description a place in psychological theories; thus Kitcher maintains that methodological solipsism is inconsistent with Marr's explanatory framework for computational theories.

Kitcher is correct in her claim that the computational level of Marr's theory is shaped by facts about the environment. But as we have just seen, while the choice of theory is affected in this way, the explanation the resulting theory gives of the computational *behavior* of the system appeals only to nonsemantic properties of representations. Like Burge, Kitcher 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.

5. Summary. Let us briefly summarize the main conclusions of this work. In section 2 we saw that Burge's argument against individualism identifies two important features of Marr's theory: (*A*) According to Marr, the veridicality of visual perception depends on contingent facts about the physical world; and (*B*) the theory is designed to explain the success of visual perception, and hence representational content is assigned in such a way that representations are true in the normal environment. I conclude that prima facie Burge's argument is sound; that is, for the reasons he cites, there can be explanatory reasons for assigning different computational interpretations to physiologically identical systems in different physical environments. Fodor's argument against this is based on an incorrect understanding of the nature of computational theories. According to Fodor, representational content must be restricted to *causal* properties of computational states; but the correct point is that semantic properties may enter into the choice of computational interpretations. However, there is a danger in placing undue emphasis on the issue of individualism, and here it is important to see that methodological solipsism captures a crucial constraint on computational theories: While there are distinct computational descriptions of the same physical system that differ only in the interpretations assigned to its computational states, *within a theory* there can be no differences in content for which there is not a corresponding difference in neural states. Fodor generates unnecessary opposition to this point by extending methodological solipsism to a restriction on the explanatory goals of computational psychology generally. The criticisms of methodological solipsism attack this restriction, rather than the real point of the principle as it is applied as a constraint on computational theories. And a proper understanding of the constraints that operate on successful computational theories is an important ingredient in grasping their explanatory power.

REFERENCES

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

Egan, F. (1991), "Must Psychology Be Individualistic?" *The Philosophical Review 100*: 179–203.

Fodor, J. A. (1981), "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology". Reprinted in *Representations: Philosophical Essays on the Foundations of Cognitive Science*. Cambridge, MA: MIT Press, pp. 225–256.

———. (1987), *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: MIT Press.

Fodor, J. A. and Pylyshyn, Z. W. (1981), "How Direct is Visual Perception? Some Reflections on Gibson's 'Ecological Approach' ", *Cognition 9*: 139–196.

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

Marr, D. (1982), *Vision*. San Francisco: Freeman.

Putnam, H. (1975), "The Meaning of 'Meaning' ", in *Philosophical Papers*. Vol. 2, *Mind, Language and Reality*. Cambridge, England: Cambridge University Press, pp. 215–271.

Stabler, E. P., Jr. (1987), "Kripke on Functionalism and Automata", *Synthese 70*: 1–22.

Ullman, S. (1979), *The Interpretation of Visual Motion*. Cambridge, MA: MIT Press.