The British Society for the Philosophy of Science

Externalism and Marr's Theory of Vision

ROBERT M. FRANCESCOTTI

I INTRODUCTION

The internalism/externalism debate has been a major issue in the philosophy of psychology lately. This dispute concerns whether mental content should be considered solely as a function of what goes on within the organism, or whether it should be considered at least partly a function of facts about the organism's external environment. Quite apart from the issue of how *folk psychology* individuates mental content, there is the question as to how the *science of psychology* should individuate mental content. It is this latter issue that will be the focus of discussion here.

   According to the internalist, psychology should seek to individuate mental content 'narrowly'. That is, it should regard content as completely determined by the organism's inner constitution. On this view, the basic tenets of any psychological theory dealing with mental content should be such that they preclude the possibility of an organism and its doppleganger (*i.e.*, its molecule for molecule duplicate) having different contentful states. According to externalism, on the other hand, psychology should appeal to a 'wide' individuation of content. That is, it should regard the content of at least some mental states as underdetermined by the organism's inner constitution, thereby leaving open the possibility of an organism and its doppleganger having different contentful states.

   Tyler Burge [1986] suggests that we should approach this issue, not as armchair philosophers guided by certain a priori, metaphysical considerations, but rather by looking and seeing what actually goes on in psychology. This seems to be good advice. What better way is there to determine how psychology should individuate content than by directing our attention towards the science itself and seeing which method of individuation makes for a more fruitful theory?

Recent writers on this issue have taken just the advice that Burge offers, and have cited the success of certain psychological theories in support of their position. One theory that is frequently cited by proponents of externalism is David Marr's [1982] theory of human vision. Marr's theory has been considered by such competent philosophers as Tyler Burge [1986], Patricia Kitcher [1988] and Robert Van Gulick [1988] to be a prime example of a respectable, and seemingly fruitful, psychological theory that's committed to an externalist characterization of mental content.

The purpose of this paper is to show that this is the wrong interpretation of Marr's theory. Of course, one cannot refute externalism by showing that a particular psychological theory is not committed to it, for there may be other, more fruitful, theories that individuate content widely. Yet, what emerges from the following discussion is that externalism is not nearly so easy to establish as one might initially think. In what follows, I shall bring to the fore the forest of complex issues underlying this superficially straightforward debate.

2 MARR'S METHODOLOGY

Marr's theory of human vision embodies three distinct levels of explanation—the computational level, the algorithmic level and the level of hardware implementation. The **computational level** is the most abstract of the three. As Marr explains, it is the level concerned with the goal of the computation, why it is appropriate, and . . . the logic of the strategy by which it can be carried out' ([1982], p. 25). To illustrate, he uses the example of the ordinary cash register (pp. 22–3). Addition is the primary function for which the cash register is designed. Hence, a computational explanation will involve a characterization of the fundamental properties of the addition function such as associativity and commutativity. Similarly, a computational explanation of human vision will involve a characterization of the information-processing tasks that the human visual system is designed to perform. The function of the human visual system is to generate veridical representations of a certain class of environmental features. Hence, the computational explanation will include a specification of the environmental features detected and an account of the strategy whereby the visual system goes from this environmental input to veridical represen-tations.

However, as Marr notes, computational explanation will involve not only a characterization of the function that is being computed, but also a specification of certain 'constraints' by which the system must abide in order to successfully and reliably compute the function (pp. 22–3). Consider again the case of the cash register. If our computational account is to adequately explain how the machine is able to carry out the addition function successfully and reliably, it will need to include a specification of the following sort of constraints.

(a) The order in which the goods are presented to the cashier should not affect the total (Commutativity).
(b) Arranging the goods into two piles and paying for each pile separately should not affect the total amount to be paid (Associativity).
(c) When an item is bought and then returned for a refund, the total expenditure should be zero (Rule for Inverses).

Likewise, in the case of human vision, computational explanation will involve the specification of certain facts about a human's relation to his external environment which are powerful enough to ensure that the visual system successfully and reliably performs its designed function—namely, the production of veridical visual representations.

Explanation at the **algorithmic level** will involve an account of the rules by which the function characterized computationally is carried out. More precisely, it will involve (i) a characterization of the representations employed by the system, and (ii) a specification of the rules by which these representations are transformed so that the function is successfully computed. As Marr illustrates (p. 24), in giving an algorithmic account of the process of addition, we might choose Arabic numerals for the representations, and we might specify rules about adding the least significant digits first and 'carrying' if the sum exceeds nine. Likewise, an algorithmic account of human vision will involve a characterization of the mental representations used (a characterization of both their syntactic and semantic properties), and a specification of the rules by which these representations are transformed as the system goes from environmental input to veridical, visual representations.

The lowest level of explanation is the level of **hardware implementation**. Explanation at this level will involve an account of how the processes described at the other two levels are physically realized. In the case of human vision, the account will be in terms of the relevant neurophysiological mechanisms and processes.

Even without going too deeply into the specifics of Marr's theory, we can already see that the theory's basic methodology commits it to explicit reference to features of the external environment. It does so in three distinct, yet closely related, ways. (These correspond quite closely to three points that Burge discusses ([1986], pp. 29–34).

First, some of the main questions that the theory is designed to answer are clearly intentional in character. And among these, many will have to do with the representation of external objects and features. The main goal of Marr's theory is to explain how humans visually interact with their environment in the way that they do. Thus, a specification of which external features are ultimately detected by the visual system will inevitably figure in the theory. Yet, in addition to telling us which external features are successfully and reliably detected, the theory will also need to explain how those features are internally represented. As noted above, an algorithmic account of the visual

process will involve a characterization of the semantic features of the representations employed by the visual system. And since some of these representations will be 'object-oriented' (*i.e.*, will be about objects in the external environment), algorithmic explanation will involve reference to the external world.

This brings us to our first main point concerning Marr's theory:

> (1) Insofar as some of the questions that the theory is designed to answer are intentional in character, concerning the representation of objective features, the theory will inevitably make reference to the external environment.

Another way in which Marr's theory makes reference to the external environment has to do with the role of constraints. As noted above, the task of computational explanation is to provide an abstract characterization of the function which takes the organism from environmental input to visual representations that are for the most part veridical. But to ensure that the visual system carries out this function successfully and reliably, certain general facts about the environment must be called upon as constraints on the processes involved. Without such constraints, there is no way to guarantee that the visual system will successfully and reliably generate veridical representations.

Consider the following example.[1] Among the various types of representations that figure in Marr's theory is a type called the 'image' ([1982], pp. 41–98). The image is the representational type that occurs earliest in the visual process, and its primary function is to carry information regarding changes in light intensity. In the construction of the image, there are channels or filters that are sensitive to spatial distributions of light intensity. They are grouped into different sizes depending upon the bands of spatial frequency to which they are sensitive, and their primary function is the detection of sudden changes in light intensity ('zero-crossings', as Marr calls them[2]) within their respective frequency bands. The outputs of these different-sized channels are what serve to generate the image. The image, in turn, forms the basis upon which later visual processes occur, processes which culminate in the generation of object-oriented representations.

At this point, Marr is faced with the task of explaining why it is that the outputs of the different-sized channels combine so as to indicate light intensity changes in the image which correspond to particular external features—*i.e.*, edges, contours or shadows in the environment. Only then can he explain how these early visual processes are able to give rise to veridical, object-oriented

---

[1] This example is also discussed by Burge ([1986], pp. 30–1).
[2] Put very generally, a zero-crossing is the place where the value of a function passes from positive to negative. In the case of a light intensity function, a zero-crossing indicates a sudden intensity change.

representations. Here Marr appeals to what he calls **the constraint of spatial localization**:

> things in the world that give rise to intensity changes in the image are (1) illumination changes, which include shadows, visible light sources and illumination gradients; (2) changes in the orientation or distance from the viewer of the visible surfaces; and (3) changes in surface reflectance (p. 68).

As Marr notes, the crucial point here is that objects or events in the world which give rise to intensity changes in the image (such as shadows, contours, light sources, creases and surface boundaries) are spatially localized, rather than scattered or made up of waves. This explains why a coincidence between the zero-crossings from different frequency bands should correspond to some single external feature (*e.g.*, a real edge, contour or shadow in the environment).

This is just one example, among many, where Marr appeals to constraints— general, contingent facts concerning how the organism relates to its environment—in order to explain how the visual system can successfully and reliably carry out its information-processing tasks. This brings us to our second main point concerning Marr's theory.

> (2) Inasmuch as the theory appeals to constraints to explain the success and reliability of the human visual system, it makes reference to facts about the environment in which the system operates.

A final way in which the theory makes reference to the external environment has to do with what Burge calls its 'success-oriented' nature ([1986], pp. 32–4). As Burge notes, Marr's theory is success-oriented in that it is designed to explain how the visual system is able to reliably generate veridical representations of various environmental features. The success-oriented nature of the theory is clearly manifested in its appeal to constraints. But it is success-oriented in another, perhaps more basic, way. It is an essential part of Marr's basic methodology that facts about the manner in which humans visually relate to their environment—*e.g.*, facts about which environmental features they can reliably detect, and facts about the environmental conditions that obtain when successful and reliable detection takes place—strongly guide the way in which the content of the representations posited is specified.

Consider, for example, what Marr says about the visual representation of three-dimensionality (pp. 215–25). Given that we sometimes interpret silhouettes as three-dimensional, Marr proposes that a 'generalized cone' (the surface generated by moving a cross-section along an axis) is a central primitive of the 3-D Model Representation (the first object-oriented representation of the visual process). He argues for such a primitive by appealing to the following three physical constraints: (1) 'each line of sight from the viewer to the object grazes the object's surface at exactly one point' (p. 219), (2) 'points

that appear close together in the image are actually close together on the object's surface' (p. 220), and (3) 'the contour generator [the set of points on the object's surface that projects to the boundary of the silhouette in the image] is planar' (p. 222)—*i.e.*, lies entirely in a single plane. Having laid down these constraints, Marr advances the following theorem: 'if the surface is smooth . . . and if restrictions 1 through 3 hold for all distant viewing positions in any one plane . . . , then the viewed surface is a generalized cone' (p. 223). Marr infers, from the constraints and the theorem, that representations of generalized cones are likely to be involved in our visual representations of 3-D objects.[3]

This example clearly shows that in characterizing the content of the representations posited, Marr is guided by facts concerning how the visual system relates to features of its environment. This brings us to our third main point concerning Marr's theory.

> (3) In characterizing the content of the representations posited and explaining how the individual representations are semantically related to one another, Marr's theory takes into consideration certain general facts about the way humans visually relate to their external environment—*e.g.*, facts concerning which environmental features they can successfully and reliably detect, and facts concerning the environmental conditions that obtain when successful and reliable detection takes place.

The fact that a psychological theory makes reference to the environment does not, in itself, imply a commitment to externalism. But the contention of Burge, Kitcher and others is that *the manner in which* Marr's theory refers to the environment does carry such a commitment. I disagree. In what follows, I shall examine more closely the three points mentioned above, and show that nothing externalist need follow. The interpretation proposed is one that remains entirely neutral on the issue.

3 A NEUTRAL INTERPRETATION

As noted in point (1) above, Marr's theory is designed to answer questions that are fundamentally intentional in character, many of which have to do with the representation of external objects and events. To see that this need not involve an externalist characterization of content, let us begin by considering Fodor's principle of methodological solipsism.

According to Fodor's [1981] original formulation of the principle, psychological inquiry should adhere to the 'formality condition' according to which 'mental states can be (type) distinct only if the representations which constitute their objects are formally distinct' (p. 277). The point of the formality condition is that the content of mental representations must

---

[3] Burge discusses this example as well ([1986], pp. 33–4).

supervene on their syntactic properties which in turn must supervene on the intrinsic properties (especially causal properties) of our neurophysiology. Adhering to this constraint does not preclude reference to the content of the states in question: it only precludes the possibility of two syntactically indistinguishable states differing in content. Indeed, far from ignoring content, the sort of theory that Fodor envisages is one which seeks to explain how content (in particular, the representation of environmental features) is a function of formal properties. As he remarks, 'By . . . exploiting the notions of content and computation *together*, a cognitive theory seeks to connect the *intentional* properties of mental states with their *causal* properties vis-á-vis behavior' (p. 241). Thus, the theory envisaged makes reference to environmental features while remaining internalist.

To take a more striking example, consider the case of a psychological theory which *type-identifies* mental states, not with the syntactic features of brain states, but with the brain states themselves. Such a theory may very well be concerned with questions that are intentional in character, questions having to do with what features of the environment are successfully and reliably detected. However, since the theory type-identifies mental states with brain states, it will obviously be a theory that is committed to local mind-brain supervenience and, hence, be internalist.

Point (2) is a little more challenging. As we recall, one of the main aims of Marr's theory is to explain how the visual system is able to generate veridical visual representations. But constraints will be an essential ingredient in such an explanation, constraints which are powerful enough to ensure that the system's information-processing tasks are performed successfully and reliably. In appealing to constraints, the theory will thereby appeal to facts about the environment within which the system operates. However, as the following example shows, this need not involve a commitment to externalism.

For ease of exposition, let us consider again a psychological theory which seeks to type-identify mental states with brain states. While such a theory is externalist, it will certainly recognize that the human visual system reliably forms veridical representations in some environments, but not in others (*e.g.*, not when the system is operating in total darkness or when the visual cortex is being artificially stimulated by electrodes). Assuming that the theory adopts an information-processing approach to the study of vision, it seeks to determine the conditions that must obtain for successful and reliable computation to take place. Specification of those conditions will, as in Marr's theory, take the form of a list of constraints that have to hold between the organism and its surroundings. For example, an account of the visual detection of three-dimensionality would not be complete unless it explained the environmental conditions that must obtain in order for veridical representations to be generated.

Thus, even for a psychological theory which endorses local mind-brain

supervenience, questions such as 'What is the logic of the strategy by which the organism arrives at its representations given the appropriate environmental input?' and 'What constraints must be satisfied in order that the function is successfully and reliably computed?' would be very appropriate. Indeed, it is the essence of an information-processing approach to address such questions—regardless of the nature of the psycho-physical identifications involved.

Perhaps the most compelling reason for thinking that Marr's theory is externalist is provided by point (3) above. Facts concerning the system's environment strongly guides the way in which Marr specifies the content of the representations posited. As noted above, Marr argues that a 'generalized cone' is one of the main primitives of the 3-D Model Representation by taking into consideration facts such as (a) each line of sight from the viewer to the object grazes the object's surface at exactly one point and (b) points close together in the image are actually close together on the object's surface, facts which are in part about the environment within which the visual system operates. As Burge contends, this implies that

> if the theory were confronted with a species of organism reliably and successfully interacting with a different set of objective visible properties, the representational types that the theory would attribute to the organism would be different, regardless of whether the individual organism's physical mechanisms were different. (Burge [1986], p. 34)

and, therefore, that the theory is externalist.

I think that Burge is mistaken. An important, yet subtle, distinction needs to be made here. We need to distinguish between

> (i) the question of whether a theory appeals to some fact X in trying to describe the nature of some theoretical entity Y,

and

> (ii) the question of whether the theory makes Y's type-identity partly a function of X.

In particular, we need to distinguish between

> (i') the question of whether a cognitive theory appeals to facts about the organism's external environment in its attempt to posit mental representations and describe their content,

and

> (ii') the question of whether the theory makes the type-identity of those mental representations partly a function of those external features.

The following story will serve to clarify this distinction.

Imagine that a team of cognitive psychologists set out to study a group of previously unstudied organisms which seem to visually interact with their environment quite successfully. These psychologists are hard-core realists

regarding mental content: before they even begin to construct their theory, they take it that the perceptual states of the organisms in question have a fairly determinate content, a content which is what it is independently of any existing theory. It is now their goal to determine just what this content really is. Suppose, moreover, that they endorse a local mind-brain supervenience. Even prior to the investigation, they believe that whatever visual states their inquiry happens to reveal, their content will be completely determined by the intrinsic properties of the brain. Finally, imagine that our psychologists are of the opinion that the best way to explain perceptual processes is to describe their role in enabling the organism to successfully interact with its environment. They, therefore, decide to adopt the methodology of David Marr: they decide to treat the organisms as information-processors, characterizing their visual states in terms of how those states enable them to successfully and reliably process information regarding their surroundings.

Burge will claim that this story is incoherent. I disagree. In order to even get started, our psychologists must first get clear on just which environmental features are reliably detected. They must, for example, determine whether the organisms can reliably detect the presence of both horizontal and vertical edges, whether they can successfully distinguish between the edges of objects and the shadows which those objects project, whether they can successfully make figure-ground discriminations, whether they can see objects as three-dimensional, and so on. Only then can they begin to computationally characterize the relevant visual processes.

Thus, despite their internalist inclinations, facts about relations to the environment will play a crucial part in guiding the theoretical inquiry or our imagined psychologists. Such facts will not only influence the sort of mental representations they decide to posit, but also influence the way they decide to characterize the content of those representations. For example, if the organisms were found to be very poor at visually detecting three-dimensionality, if behavioral observation revealed that the world looks basically 'flat' to them, then Marr's 'generalized cones' would probably not figure anywhere in the specification of their mental content. Or if behavioral observation were to reveal that these organisms are much better than humans at detecting changes in light intensity, then the description of their early visual processes are likely to be far more complex than what is found in Marr's description of the image—perhaps involving a greater number, and variety, of channels.

Despite their internalist inclinations, our psychologists will also need to consider the environmental conditions that obtain when successful and reliable detection takes place. Suppose, for example, that our imagined creatures can visually detect three-dimensionality, but only under certain very special conditions (only when, say, the object viewed lies at a certain angle with respect to their line of sight, or only when the light intensity changes on the object's surface fall within a certain range). Such facts would weigh quite

heavily in deciding how the representations concerning three-dimensionality are generated from, and semantically related to, representations occurring earlier in the visual process. And with such considerations in mind, it is quite likely that our psychologists would be led to give an account of the representation of three-dimensionality that is quite different from the account that figures in Marr's theory.

The theory envisaged appeals to facts concerning the organism's environment in its attempt to posit mental representations and describe their content, but it does not allow that the type-identity of those representations be a function of those environmental features. Thus, there is no contradiction in supposing that point (3) above applies to an internalist theory.

Nor is there any contradiction in supposing that the environmental features considered will figure in their final theory. Given that they take an information-processing approach to the study of vision, they will, like Marr, wish to characterize the psychological states in terms of how they enable the organism to successfully interact with its environment, how they play a part in allowing the organism to extract useful environmental information. Environmental considerations will, therefore, not only guide them in their inquiry, they will also be referred to in their final theory, regardless of the fact that they advocate local mind-brain supervenience.

Before closing, I would like to forestall an anticipated objection to my neutral interpretation.

4 AN APPEAL TO THE NOTION OF INFORMATION

Burge may object that I have failed to do justice to the 'success-oriented' nature of Marr's theory. I anticipate the following line of objection. Just as the notion of knowledge carries with it the notion of success (one cannot be said to know that p, for some proposition p, unless one's belief that p is true), the notion of *information* also carries with it the notion of success. While it does make sense to say that an organism might possess misinformation regarding its environment, the notion of misinformation makes sense only when it is an exception and not the rule. If an organism is to be considered a genuine information-processor vis-à-vis some external feature X, then the organism must be a reliable detector of the presence of X. If it typically fails at detecting, say, the presence of horizontal edges, then it cannot be said to be an information-processor vis-à-vis horizontal edges. It seems to be part of the very concept of information that an organism can only be said to carry and process information about its environment if it 'gets things right' most of the time.

But, the argument continues, Marr's methodology presupposes that the human visual system is an information-processor. Hence, the theory presupposes that the visual system acquires, at least for the most part, veridical representations. This precludes the possibility of a human visually misrepre-

sentating his environment most of the time. Thus, if my physical duplicate were situated in an environment that is radically different from mine (an environment in which most of my current visual representations would be non-veridical), then we cannot say that his representations are mostly non-veridical: instead, we must say that his representations are different from mine. This amounts to a denial of local mind-brain supervenience. Hence, Marr's theory is externalist.

This argument appeals to a sort of principle of charity: we should interpret the content of psychological states in such a way that the organism ends up having mostly veridical representations. Yet, while Marr's theory does presuppose a sort of principle of charity, the principle envoked is not quite so strong. Marr is treating the human visual system as an information-processor. That is because he is trying to describe how the human visual system *standardly* operates, and the typical human does generate mainly veridical representations. However, this does not preclude the possibility of a human visually misrepresenting his environment most of the time. It only denies that the *typical* human visual system might misrepresent its environment most of the time. It is quite consistent with Marr's approach that there be a human who, as a result of being situated in an environment to which he is visually ill-adapted, happens to generate mostly non-veridical representations. Indeed, Marr's approach not only allows this possibility, but it provides a means for explaining it in terms of deviation from normal visual functioning. Thus, while Marr's methodology presupposes the human visual system standardly operates so as to generate mostly veridical representations, this principle of charity is not the sort that defies local mind-brain supervenience.

5 CONCLUDING REMARKS

While environmental considerations play a large role in Marr's theory, the role played need not involve a wide-individuation of content. Of course, there remains the deeper question as to whether psychological inquiry *is at its best* when it regards the type-identity of representations as underdetermined by inner constitution. However, given that Marr's theory of vision is one of the more fruitful theories on the market today, we have given the externalist something to worry about.[4]

Syracuse University

---

[4] Thanks to Robert Van Gulick, John Hawthorne and Melissa Joulwan for helpful comments on an earlier draft of this paper.

Robert M. Francescotti

REFERENCES

BURGE, T. [1986]: 'Individualism and Psychology', *Philosophical Review*, **95**, pp. 3–45.
FODOR, J. [1981]: 'Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology', in *Representations* MIT Press, pp. 225–53.
KITCHER, P. [1988]: 'Marr's Computational Theory of Vision', *Philosophy of Science*, **55**, pp. 1–24.
MARR, D. [1982]: *Vision*. New York, W. H. Freeman and Co.
VAN GULICK, R. [1989]: 'Metaphysical Arguments for Internalism, and Why They Don't Work', in S. Silvers (*ed.*) *Re: Representation*. Dordrecht, Kluwer, pp. 151–60.