



**OXFORD JOURNALS**  
OXFORD UNIVERSITY PRESS

## Mind Association

---

Individualism, Computation, and Perceptual Content

Author(s): Frances Egan

Source: *Mind*, New Series, Vol. 101, No. 403 (Jul., 1992), pp. 443-459

Published by: [Oxford University Press](#) on behalf of the [Mind Association](#)

Stable URL: <http://www.jstor.org/stable/2253897>

Accessed: 17/02/2011 14:58

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=oup>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).



Oxford University Press and Mind Association are collaborating with JSTOR to digitize, preserve and extend access to *Mind*.

<http://www.jstor.org>

# *Individualism, Computation, and Perceptual Content*

FRANCES EGAN

## *1. Introduction*

Individualism in psychology is a thesis about how mental states are to be taxonomized. As Tyler Burge characterizes it, individualism is the view that

the mental natures of all a person's or animal's mental states (and events) are such that there is no deep individuating relation between the individual's being in states of those kinds and the nature of the individual's physical or social environments. (1986, pp. 3-4)

Individualism has sometimes been formulated as a supervenience thesis, according to which psychological states are said to supervene on intrinsic, physical states of the organism to which they are ascribed (e.g. Stich 1983, pp. 164-5). Any differences between organisms not reflected in their intrinsic physical states are not psychologically relevant, it is claimed, and should be ignored by psychological theory. Thus, according to individualism, I and my Twin Earth counterpart are psychologically identical, in virtue of the fact that we are, according to Putnam's story, molecule for molecule identical.

In a series of important papers (especially 1979 and 1986), Tyler Burge has argued that individualism is false with respect to a wide range of explanatory kinds in psychology, including the intentional states invoked in folk psychological explanation. Most strikingly, perhaps, Burge and others have claimed that perceptual theories, including David Marr's computational theory of vision, individuate perceptual states in part by reference to the environment of the subject possessing them, and so violate individualism (e.g. Burge 1986 and 1988, Kitcher 1988, and Davies 1991).

As formulated above, individualism says nothing about the *contents* of mental states. While individualism is generally construed as a thesis about the individuation of the propositional attitudes, that is, beliefs and desires—states which have contents—nothing in the above formulations requires that the mental states so individuated must have propositional contents.<sup>1</sup> In fact, proponents *and* opponents of individualism (in psychology) often cite analogies from other sciences—in which the relevant states are claimed either to supervene on local physical features or to fail to do so—in support of their views (e.g. Burge 1986, Fodor 1987, ch.

<sup>1</sup> This is also true of Fodor's characterization of *methodological individualism*, "the doctrine that psychological states are individuated *with respect to their causal powers*" (1987, p. 42). Fodor's characterization of individualism differs in important respects from both Burge's and Stich's. (The latter two I take to be roughly equivalent.)

2). These arguments presuppose that the nature of the case is not significantly changed by the fact that psychological states typically have propositional contents.

Nevertheless, the usage has not been consistent. Individualism is sometimes characterized as a thesis about how mental *contents* are to be individuated. I propose, however, to use the term “internalism” for the view that holds that the *contents* of mental states supervene on intrinsic physical states of the subject, and hence are individuated “narrowly”, without essential reference to the subject’s physical and social environment. The view that denies that mental contents supervene on intrinsic physical states of the subject, claiming that they are individuated in part by reference to the subject’s environmental or social context, hence “widely”, I shall call “externalism”.

It is generally assumed that intentional mental states are individuated by their contents, that they have their contents *essentially*. Proponents of individualism, unless they are eliminativist about intentional contents, have therefore found themselves burdened with the task of articulating and defending some notion of *narrow* content, where by “narrow” we mean *supervening locally on the subject*. Skepticism about the possibility of an adequate account of narrow content has led others to embrace anti-individualism, on the assumption that if the *content* of mental states fails to supervene on intrinsic states of the individual subject, then the mental states possessing such content must also fail to supervene on intrinsic states of the individual subject.

In this paper I wish to challenge the widely held assumption that individualism entails either content internalism or content eliminativism,<sup>2</sup> while content externalism entails the falsity of individualism. I shall confine my attention to computational psychology. I shall argue, first, that computational theories are individualistic—they taxonomize mental states without essential reference to the subject’s environment. Representational contents, I suggest, play a role in computational psychology analogous to the role played by models in the physical sciences. I shall then argue that the content attributed to computational states in the explanatory models of computational theories of perception are, typically, *wide*, illustrating my argument by reference to Marr’s theory of vision. Therefore, contrary to the above assumption, computational theories of perception are both individualistic *and* externalist. This is possible only because mental states, as characterized by computational psychology, do not have their contents essentially.

## 2. Computation and individuation

My argument that computational theories are individualistic depends in part upon the view that the goal of such theories is to characterize the mechanisms under-

<sup>2</sup> Content eliminativism either denies that mental states have content or denies that content plays a genuine explanatory role in psychology: see e.g. Churchland (1981) and Stich (1983).

lying our various cognitive capacities, and further, that this goal is best served by theories which taxonomize states individualistically.

Cognitive psychological theories aim to characterize human cognitive processes. *Computational* cognitive theories construe human cognitive processes as a species of information-processing, and the systems that subservise such processing as symbol manipulating systems. This is to imply that some of the events postulated within the system can be consistently interpreted as having a meaning in a certain domain.

My claim that the goal of computational psychological theories is to characterize the *mechanisms* underlying our cognitive capacities may appear to violate a widely accepted principle about the proper approach to the study of information-processing systems. As David Marr (1982) and others (e.g. Ullman 1979) have cogently argued, an information-processing system should be analyzed at several distinct levels of theory. Most importantly, it has been claimed, the nature of the information-processing task itself—the computation performed by the system—needs to be understood independently of any attempt to characterize the physical mechanisms supporting the computation. According to Marr:

If one believes that the aim of information-processing studies is to formulate and understand particular information-processing problems, then it is the structure of those problems that is central, not the mechanisms through which they are implemented. (1981, pp. 139-40)

Marr criticizes Newell and Simon's work on production systems because, he claims, it amounts to "studying a mechanism, not a problem":

The mechanisms that such research is trying to penetrate will be unraveled by studying problems, just as vision research is progressing because it is the problem of vision that is being attacked, not neural visual mechanisms. (1981, p. 140)

But to claim, as I do, that the goal of computational theories is to characterize the mechanisms underlying cognitive capacities is not to specify the level of abstraction at which, or the vocabulary in which, the mechanisms should be described. In particular, it is not to maintain that computational theories aim to characterize cognitive mechanisms only at what Marr calls the third level of description—the specification of the neural circuitry which implements the computation. The claim that cognitive scientists should study the information-processing problem independently of the physical mechanisms implementing the process can plausibly be construed as a recipe for achieving the correct characterization of cognitive mechanisms. On such a construal, my claim that the goal of computational psychology is to characterize the mechanisms underlying cognitive capacities is consistent with the expectation that the complete characterization of such mechanisms will include components corresponding to Marr's three levels of analysis—a specification of *the function computed by the cognitive mechanism*, a description of *how the function is computed* (i.e. a specification of a representation and algorithm), and a description of the neural hardware supporting the computation.<sup>3</sup>

<sup>3</sup> Marr, somewhat misleadingly, called the specification of the function computed by

In treating human cognitive processes as a species of information-processing, and the systems that subserve such processing as symbol manipulating systems, computational theories construe cognitive processes as formal operations defined over symbol structures. To describe something as a symbol is to imply that it is semantically interpretable, but (and this is the important point) its type identity as a symbol is independent of any particular semantic interpretation it might have. Symbols are just functionally characterized objects whose individuation conditions are specified by a realization function  $f_R$  which maps equivalence classes of physical features of a system to what we might call “symbolic” features. Formal operations are just those physical operations that are differentially sensitive to the aspects of symbolic expressions that under the realization function  $f_R$  are specified as symbolic features. The mapping  $f_R$  allows a causal sequence of physical state transitions to be interpreted as a *computation*.

Given this method of individuating computational states, two systems performing the same operations over the same symbol structures are computationally indistinguishable. If two systems are physically identical, then they serve as domains for the same class of realization functions. Consequently, there can be no *computationally relevant* grounds for attributing a particular symbolic property to one that would not be grounds for attributing it to the other. So if two systems are molecular duplicates then they are computational duplicates.<sup>4</sup> Computational descriptions are individualistic: they type-individuate states without reference to the subject’s environment or social context.

Actually, a stronger conclusion can be drawn. To the extent that computational processes are construed as *modular* processes,<sup>5</sup> even the *internal* environment is irrelevant to the type-individuation of the computational states of a system. To use an example of Martin Davies’ (which he employs in an argument *against* individualism in Davies 1991), imagine a component (module) of the visual system, called the *visex*, which computes, say, a representation of features of the surface structure of an object based on information about binocular disparity. Now imagine that within the auditory system of some actual or imagined creature there is a component that is physically identical to the *visex*. Call this component the *audex*. According to theories of auditory processing, the *audex* computes a rep-

---

the system the “theory of the computation”. I shall follow Marr in using this terminology to refer to Marr’s topmost level of analysis. What I call a “computational theory” (following standard usage) comprises all three levels of analysis.

<sup>4</sup> Cummins makes this point in his 1989, p. 81. However, he concludes from this that representational content is shared by computational duplicates. I criticize this argument below.

<sup>5</sup> David Marr says;

Computer scientists call the separate pieces of a process its *modules*, and the idea that a large computation can be split up and implemented as a collection of parts that are as nearly independent of one another as the overall task allows, is so important that I was moved to elevate it to a principle, the *principle of modular design*. (1982, p. 102)

Marr’s own theory of early vision respects the principle of modular design.

resentation of certain sonic properties. We can imagine a particular visex and audex removed from their normal embeddings in visual and auditory systems respectively and switched. Since the two components are by hypothesis physically identical, the switch should make no discernible difference to the behaviour of the creatures, nor to their internal goings-on. The two components are computationally identical, despite the difference in their normal internal environments.

It might be objected that I have construed the goal of computational psychological theories too narrowly. Characterizing the mechanisms underlying human cognitive capacities is one goal of computational psychological theories, but computational theories are also typically concerned to explain the contribution of cognitive mechanisms toward the overall success of the organism in its environment. Accordingly, it is claimed, we would expect such theories to individuate psychological states with an eye to how the mechanisms are embedded in larger systems within the organism and ultimately in the organism's normal environmental niche—that is, we would expect them to individuate psychological states non-individualistically.<sup>6</sup>

It is reasonable to assume that psychological theories will contribute to a satisfactory account of organism/environment interaction that would explain an organism's success, or failure, in its normal environment. But, I maintain, the explanation of organism/environment interaction is not the primary goal of computational theorizing, and such explanations are forthcoming only when a computational theory is supplemented by further assumptions about the normal environment in which the described cognitive mechanisms are deployed. In fact, precisely because the hypothesized cognitive mechanisms are assumed to be invariant across environmental changes, we can see why *this* mechanism would not have been adaptive had the environment been different, and why it might cease to be adaptive if the environment changes.

It might be objected that assumptions about the subject's environment play an essential role in the description of the cognitive mechanisms themselves, particularly the mechanisms underlying perception.<sup>7</sup> In the first place, the inputs to perceptual mechanisms are often characterized in terms of their typical environmental causes. Secondly, and perhaps more importantly, large scale assumptions about the environment are built into the mechanism itself. The solutions to information-processing problems solved by cognitive mechanisms are often underdetermined by the information contained in the inputs to the mechanism (the so-called "poverty of the stimulus" phenomenon). The processing is achieved only with the help of additional information which is assumed to be innate. The innate information available to the perceptual mechanisms concerns

<sup>6</sup> For versions of this argument see Burge (1986), Davies (1991), Dennett (1987), Kobes (1990), and van Gulick (1989). See Burge (1986) and Kitcher (1988) for detailed arguments that Marr's theory of vision fits this model. See Segal (1989, 1991) and Egan (1991) for arguments against the latter claim.

<sup>7</sup> See Burge (1986) and Kitcher (1988). The argument concerns Marr's theory of vision, but it can be generalized.

very general features of the subject's normal environment; because this information is true the subject's experience is typically veridical.<sup>8</sup>

Andy Clark warns against studying intelligent systems independently of the complex structure of their natural environments. He cites what he calls the "007 principle" as an important maxim of cognitive theorizing:

In general, evolved creatures will neither store nor process information in costly ways when they can use the structure of the environment and their operations upon it as a convenient stand-in for the information-processing operations concerned. That is, know only as much as you need to know to get the job done. (1989, p. 64)

Given the apparently indispensable role of environmental information, both for specifying the inputs to the perceptual mechanisms and for characterizing the additional (innate) information required for cognitive processing to proceed, it might be argued that the nature of the environment plays an essential role in the characterization of the cognitive mechanisms themselves. This conclusion, however, does not follow. Environmental information, while often crucial for theory construction and articulation, does not function essentially as part of the computational theory's individuating apparatus.

In the first place, while the inputs to perceptual mechanisms are often given a distal interpretation, the fact remains that the realization function  $f_R$  determines how the computational states of the system are individuated. This function individuates computational states *non-semantically*, that is, independently of any particular semantic interpretation such states may have. This point may be obscured by the fact that the realization function is not always completely specified. Some computational processes may be postulated without an explicit specification of the algorithms that carry out the processing, or even of the symbolic tokens that the missing algorithms would be defined over. In such cases, the theorist may rely on a presumptive semantic interpretation of underlying states to informally characterize the hypothesized procedures. The important point is that while the semantic interpretation does provide a useful description of what the system does, it does not serve to *individuate* the underlying computational states. Rather, it plays a *reference-fixing* role, giving us a way to refer to the underlying states, which must be presumed to be independently characterizable if the account is to be genuinely computational. A computational theory is committed to the existence of a fully specifiable formal account of the cognitive processes it attempts to characterize. Otherwise the processes are not programmable and the

<sup>8</sup> An example of such an assumption is Shimon Ullman's *rigidity assumption*, which says that "any set of elements undergoing a two-dimensional transformation has a unique interpretation as a rigid body moving in space and hence should be interpreted as such a body in motion" (1979). Ullman has proved that three distinct views of four non-coplanar points in a rigid body are sufficient to determine its three-dimensional structure. In a world like ours where most things are rigid, a process that incorporated the rigidity assumption—as Marr's structure-from-motion module is hypothesized to do—would generally be successful in recovering the three-dimensional structure of distal objects from three such views.

theory reneges on its promise to provide a purely mechanical (hence, physically realizable) account of cognitive processes.

Secondly, general assumptions about the environment are claimed by Marr and others to be incorporated and used by perceptual mechanisms in solving the information-processing tasks set them by nature; however, it does not follow that the mechanisms themselves are type-individuated by reference to the environment. The assumptions are built into the mechanism only in the following sense—the mechanism operates in such a way that if the assumptions are true it will succeed in recovering information about the environment from information in the input.<sup>9</sup> The important point is that the *same* perceptual mechanism could occur in an environment where the relevant assumptions were radically false,<sup>10</sup> although in such an environment we might expect the mechanism (and the organisms containing it) to be short-lived.

Finally, to say that a computational theory is individualistic is *not* to imply that the theorist must ignore the subject's environment in constructing her theories. The expectation that the cognitive mechanisms of evolved creatures are well-suited to the environment will constrain computational accounts. The theorist will typically exploit environmental knowledge in attempting to uncover the computational problems that need to be solved and the nature of the mechanisms required to solve them. She must not postulate mechanisms that are implausible from an environmental or evolutionary point of view. However, the mechanisms themselves, as characterized by computational theories, supervene on the physical states of the subjects possessing them.

### *3. The role of content*

Tyler Burge (1986) has argued that a defence of individualism in psychology is a two-part task. The individualist must not only show that the presuppositions of psychology are purely individualistic, she must also explicate an individualist language that would allow the attribution of narrow content to the subject's psychological states. I have attempted the first task above, arguing that computational theories are committed to individualist taxonomic principles. Others who have argued that psychology is individualistic have typically shouldered the second burden as well, attempting to articulate and defend a notion of mental content that is purely individualistic (e.g. Fodor 1987 ch. 2, Block 1986, Segal 1989 and 1991). But since computational theories individuate psychological states by non-semantic criteria, the individualist who restricts her attention to computational theories need not take on the second task. Whatever content computational states have—narrow or wide—the states themselves supervene on the intrinsic physical

<sup>9</sup> For example, the mechanism characterized by Shimon Ullman succeeds in recovering structure from motion in worlds where (most) objects are rigid.

<sup>10</sup> See Segal (1989) and Egan (1991) for detailed argument on this point.



states of the subject possessing them, and are thus shared by doppelgängers in the thought experiments.

One individualist who seems not to have fully appreciated the implications of computational individuation is Robert Cummins. He argues as follows:

The CTC [Computational Theory of Cognition] must hold that the capacities it seeks to explicate retain their identity across differences in *non-computational* factors. It must therefore cleave to the viability of a kind or aspect of content that is narrow with respect to causal and historical features not mirrored in computational architecture. (1989, p. 119)

Cummins is right that the cognitive capacities characterized by computational theories are assumed to be invariant across environmental differences. But he is mistaken in concluding that computationalists require a notion of content that supervenes on the subject's computational architecture. There is no *computationally motivated* reason why the interpretation of computational states should be narrow.<sup>11</sup>

The assumption that individualists require a notion of narrow content is typically based on a more fundamental assumption shared by most participants to the debate—that psychological states have their contents essentially.<sup>12</sup> This is perhaps understandable given the tendency to identify such states with propositional attitudes. Standard accounts of the propositional attitudes construe them as relations (to propositions or mental representations that have their meanings essentially), although such a construal is not mandatory.<sup>13</sup> In any event, the assumption that propositional attitudes have their contents essentially is perhaps plausible inasmuch as we have no other way to characterize the internal states posited by commonsense psychology as the causes of behaviour except by reference to their contents. But it is not plausible for mental states construed as computational states, because the latter are individuated by computational theories according to non-semantic criteria.

Computational theories construe cognitive processes as formal operations defined over symbol structures. To speak of these structures as symbols is to imply that the postulated structures can be consistently interpreted. Their interpretation is given by an *interpretation function*  $f_i$  that specifies a mapping between elements of symbol structures and elements of some represented domain. For example, to interpret a device as an adder involves specifying an

<sup>11</sup> Thus, Cummins dismissal of Millikan's account of mental representation, on the grounds that since it is externalist it is incompatible with the computational theory of cognition, is ill-founded.

<sup>12</sup> See e.g. Tyler Burge (1979, 1986, p. 15f). Fodor says "...I suppose (it) to be uncontentious that mental states have their contents essentially, so that typological identity of the former guarantees typological identity of the latter..." (1987). Jackson and Pettit challenge this assumption in Jackson and Pettit (1988).

It is clear from Cummins' account of computation that he does not assume that computational states have their contents essentially, despite the suggestion to the contrary in the passage quoted above.

<sup>13</sup> For an account that does not construe propositional attitudes as relations see Matthews (1990).

interpretation function  $f_i$  that pairs symbolic states of the device with numbers. A computational account can plausibly claim to have shown how a device actually cognizes only if there exists an interpretation function that maps computationally specified states to appropriate contents in a fairly *direct* way, although it is notoriously hard to specify precisely the conditions on direct interpretation.<sup>14</sup> Assuming that these conditions are met, and that the theory is technically (biologically) feasible, the computational account demonstrates how the device *could* compute the hypothesized cognitive function.

Computational theories provide a *formal* characterization of the functions computed by cognitive mechanisms. The functions are specified mathematically, and inputs and outputs of the processes that compute the functions are characterized non-semantically. The interpretation function  $f_i$  provides an *intentional* specification of a function, by characterizing the representational tokens over which the hypothesized process is defined in terms of some represented domain. It is misleading, however, to construe the content assigned by  $f_i$  as part of the computational theory itself, rather than as a feature of the expository apparatus that both renders the theory's formal exposition intelligible and allows the postulated computational process to be seen as *cognition*.

Let me elaborate this point. It is a truism that the postulates of a theory can be understood more readily when embedded in a familiar model than when given a purely formal (that is, mathematical) exposition. Maxwell's efforts to represent Faraday's lines of force by the flow of liquid through tubes was an attempt to make intelligible a purely formal exposition of unfamiliar phenomena (in this case, electromagnetic phenomena) by appeal to systems governed by the laws of mechanics, which have the status of familiar principles. In the case of mechanical models, the relevant similarity between the model and the modelled phenomena is a nomic isomorphism, that is, an isomorphism between two corresponding sets of laws. In the computational case, the interpretation function  $f_i$  that pairs symbolic expressions with contents specifies an isomorphism between computational states and features of the represented domain. The suggestion that the contents assigned to the representational structures postulated by computational theories should be understood as *models* of such theories is made plausible by the fact that an interpretation can aid our understanding of such formal accounts in two distinct respects. In the first place, while it is possible to specify the function computed by a computational device purely formally, rather than by way of an interpretation, doing so would make understanding the computational account exceedingly difficult. Secondly, given that the questions that define a psychological theory's domain are usually couched in intentional terms, an intentional specification of the postulated computational processes demonstrates that these questions are indeed answered by the theory, perhaps in conjunction with auxiliary assumptions. For example, a theory of mental arithmetic must explain how a subject is able to compute simple sums and products. It is only under an interpretation of some of the internal states of the subject as representations of num-

<sup>14</sup> For discussion of this issue see Cummins (1989, ch. 8).

bers that the computational processes postulated by the theory are revealed as addition and multiplication. Thus, the interpretation *explains* the computational account.

If the foregoing account is correct, then computational theories of cognition are not intentional, although they have intentional models, where an intentional model is just an interpretation that treats the device as computing a cognitive function.<sup>15</sup>

Construing interpretations as explanatory models of computational theories helps to explain other features of computational theorizing. Often when a theory is incompletely specified the study of a model of the theory can aid in the further specification of the theory itself. As previously noted, a computational theorist may resort to characterizing a computation partly by reference to features of some putative represented domain, hoping to supply the formal details (i.e. the theory) later. For example, the representational tokens postulated by a theory of visual perception might be characterized in terms of the distal interpretations that would be assigned to such states under an intuitively plausible interpretation function. In such a case, the computational theory is expressed through one of its models, and the language used to express the theory should not be construed as basic. The hypothesized function could, in principle, be described formally. The philosopher interested in uncovering the individuating principles of the computational theory should not assume that they can be read directly off the intentional specification of the functions computed by the postulated computational mechanisms. To argue that the states picked out are individuated essentially by reference to their distal interpretations (or any intentional specification) is to mistake an adventitious feature of a theory's model for an essential feature of the theory itself.<sup>16</sup>

It is often noted that, since an interpretation is just a structure-preserving mapping between symbolic elements and elements of some represented domain, the representational contents of computational states will typically be non-unique. If interpretations are understood as explanatory models, this is just what one would expect. The existence of unintended models does not undermine the explanatory usefulness of an intended interpretation. The choice of an explanatory model for a computational theory is based on extrinsic (i.e. noncomputational) considerations, the most important of which, is demonstrating that the questions that define the theory's domain are indeed answered by the theory. The fact that a hypothesized visual system could be interpreted as computing a function on the auditory domain (or as calculating the batting averages of the New York Mets) would not undermine the theorist's claim to have described a possible visual sys-

<sup>15</sup> A cognitive function can be characterized informally as a function whose arguments and values are epistemically related. Thus, the outputs of the computation can be seen as cogent or rational given the inputs.

<sup>16</sup> I argue in (1991) that Burge makes this mistake in interpreting Marr's theory of vision.

tem, assuming that the mechanism could be consistently and directly interpreted as computing the appropriate function on the visual domain.

#### 4. *Perceptual content*

I turn now to the representational contents assigned to the data structures subserving perception. I shall focus on a favourite example of philosophers—the theory of vision developed by David Marr and his colleagues. My discussion is intended to support the general account of content articulated in the previous section.

In accordance with his methodology for the explanation of an information-processing capacity, Marr's theory of vision is deployed at three distinct levels of description—the specification of *the function computed* by various visual processes hypothesized by the theory, the *algorithmic implementation* of the hypothesized functions, and *the hardware implementation* of the hypothesized algorithms. The “topmost” level—what Marr called the “theory of the computation”—is the most developed aspect of the theory.

The goal of the visual system is to derive a representation of three dimensional shape from information contained in two-dimensional images. Marr's theory divides this task into three distinct stages, each involving the construction of a representation, tokens of which serve as inputs to subsequent processes. Vision culminates in a representation that is suitable for the recognition of objects. Innate assumptions of the sort described earlier, incorporated into the visual system itself, and reflecting physical constraints on the pairing of retinal images with distal shapes, allow the postulated mechanisms underlying early vision to recover information about the distal scene based only on information contained in the image. Early visual processing is thus “data-driven”.

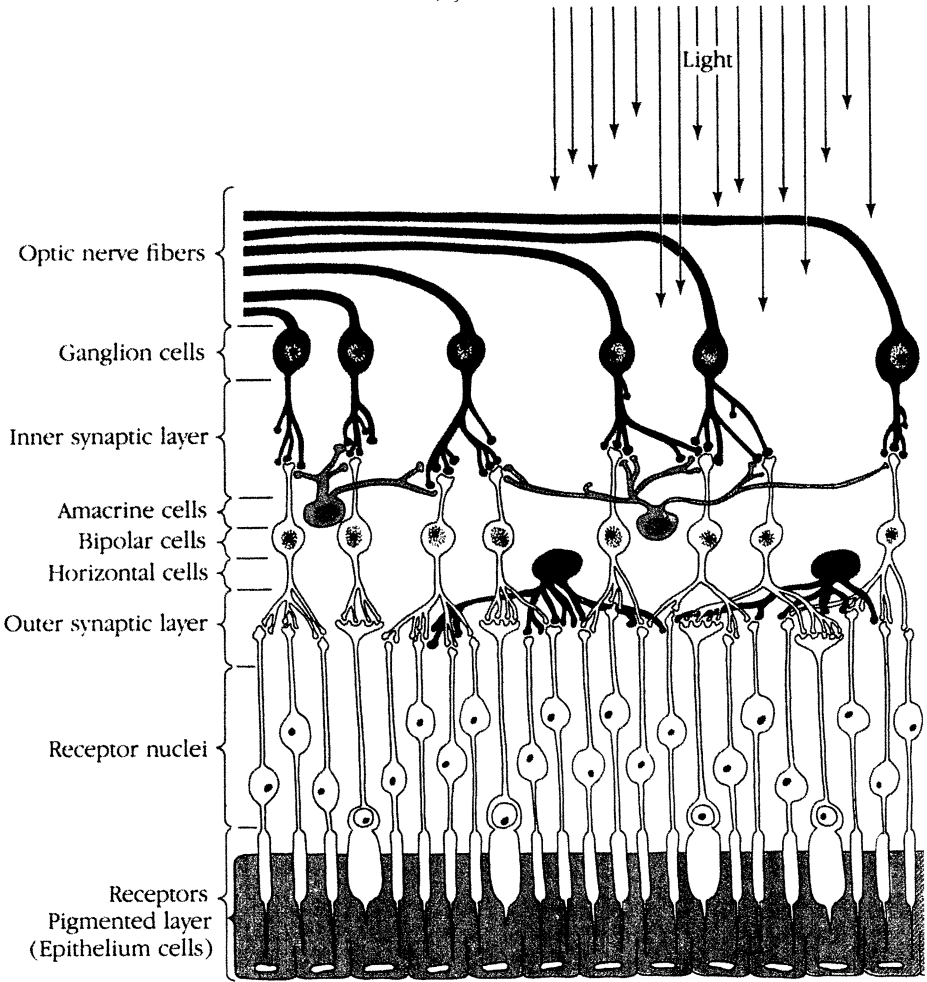
Interpreters of Marr (individualists and anti-individualists alike) have construed the theory of the computation, that is, the specification of the function computed by the visual system, as *intentional*. More specifically, it has been argued that the functions computed by the various modules of the visual system (i.e. what the system does) are individuated essentially by reference to the contents of the representational tokens that form the inputs and outputs of these modules (see Burge 1986, Kitcher 1988, Davies 1991, Segal 1989 and 1991). This claim is based on a misreading of Marr's theory, and more generally, on a misunderstanding of the computational approach to cognition. In his exposition of the theory of vision, Marr often describes the postulated visual processes in terms of features of the distal environment that typically co-vary with the representational tokens that form the inputs and outputs to the processes. This suggests that the theory has intentional *models*, in particular, that it has *externalist* models, but not that the theory is intentional. In discussing the levels-of-explanation methodology, Marr explicitly points out that the theory of the computation is a *formal* characterization of the function(s) computed by the various processing modules. The following diagram, taken from Marr's book *Vision* (1982, p. 338), depicts

(top) the mathematical formula that describes the initial filtering of the image, and (below) a cross section of the retina which implements the computation.

$$\nabla^2 G * I(x, y),$$

$$\text{where } \nabla^2 G(r) = -\frac{1}{\pi\sigma^3} \left(1 - \frac{r^2}{2\sigma^2}\right) \exp\left(\frac{-r^2}{2\sigma^2}\right)$$

(a)



(b)

The point to note is that the function computed by the retina ((a) in the diagram) is characterized formally. Marr says the following:

I have argued that from a computational point of view [the retina] signals  $\nabla^2 G * I$  (the  $X$  channels) and its time derivative  $\partial/\partial t (\nabla^2 G * I)$  (the  $Y$  channels). From a computational point of view, this is a precise specification of what the retina does. Of course it does a lot more—it transduces the light, allows for a huge dynamic range, has a fovea with interesting characteristics, can be moved around, and so forth. What you accept as a reasonable description of what the retina does depends on your point of view. I personally accept  $\nabla^2 G$  [(a) in the diagram] as an adequate description, though I take an unashamedly information-processing point of view. (p. 337)

$\nabla^2 G$  is a function that takes as arguments two dimensional intensity arrays  $I(x,y)$  and has as values the isotropic rates of change of intensity at points  $(x, y)$  in the array. The implementation of this function is used in Marr and Hildreth's (1980) model to detect *zero-crossings*, which correspond to sudden intensity changes in the image.<sup>17</sup> Marr grants that the formal specification of the function computed by the retina may not make what the retina does *perspicuous*. Nonetheless, from an information-processing point of view, the formal specification is "adequate".

The representational tokens over which the processes postulated in Marr's theory are defined are built up out of sets of *primitives*. The *primal sketch*, for example, is constructed out of blobs, bars, edges, and terminations. The fact that Marr calls these primitives "edges" and "bars" does not mean that they represent properties of the distal scene; indeed Marr explicitly cautions against such an interpretation. He is careful to point out that these primitives, considered individually, do not reliably co-vary with what we take to be salient features of the distal scene (for example, object boundaries) and so do not have "physical reality".<sup>18</sup> They are treated, in the theory, as uninterpreted *structures*. It is their structural properties—position, length, width, and orientation—that are computationally significant. Grouping processes operate on the primitives in virtue of their structural properties (see Marr 1982, p. 53 and pp. 71-3).

To summarize the above discussion: (1) The theory of the computation is the formal specification of the functions computed by the visual system. (2) Inputs and outputs of the hypothesized processes are characterized in the theory in terms of their computationally significant properties, that is, their structural properties. (1) and (2) support my characterization of Marr's theory as individualistic, and also my claim that the theory is not intentional.

Marr is attempting to characterize a mechanism that we know does reliably recover information about the environment, so he is concerned to find structures that correspond to real physical changes. One needs to be somewhat cautious in

<sup>17</sup> It is important to note that the mathematical formula that describes the  $\nabla^2 G$  function is not assumed by Marr and Hildreth to be explicitly represented in the retina.

<sup>18</sup> They do reliably co-vary with properties of the image. Gabriel Segal (1989) concludes that *edge* is assigned narrow content in the theory. But there would seem to be no motivation for such a move. Marr's caution against assigning distal interpretations to primitives at this stage should not be construed as entailing that they have narrow content.

drawing implications from this fact. To point out that certain data structures are reliably correlated with salient features of the distal scene is not, in itself, to attribute wide content to these structures. Nowhere does Marr do the latter, and it would certainly be wrong to attribute to him a causal covariance theory of content, indeed *any* theory of content. Nevertheless, such structures *are* candidates for distal interpretation in explanatory models of the theory. The *raw primal sketch*, which contains information from several distinct  $\nabla^2 G$  channels, is the first data structure which correlates (in *this* world) with salient physical properties, and so is the earliest candidate for distal interpretation. Subsequent data structures, notably the 2.5-D sketch and the 3-D model representation, can be consistently given distal interpretations. In some *other* environment, however, structures in the retinal image might be reliably correlated with, say, object boundaries, and so would be plausible candidates for distal interpretation in explanatory models appropriate to that world.

It would be consistent with the above that at least some representations are assigned narrow content. The problem with this suggestion is that there is no motivation for the ascription of narrow content, and no evidence that Marr has any interest in the possibility of a non-distal interpretation. The fact that structures in the primal sketch reliably co-vary with features of the image, for example, does not justify construing these structures as representing features of the image. There is, however, a clear motivation for the ascription of wide content.

Ascribing content to the postulated representational structures helps to make the formal apparatus of the theory intelligible. Ascribing *wide* content enables us to see that the visual system is able to perform the cognitive task that defines the theory's domain—it can recover 3-D shape relations among objects from information contained in two dimensional projections, *in its normal environment*. By interpreting (some of) its states as referring to aspects of the distal scene we can see how the visual system *could* perform the antecedently characterized cognitive task, and in the absence of a competing account, plausibly how it *does*. It is therefore likely that explanatory models of Marr's theory of vision are *externalist*. The content ascribed in such models does not supervene on intrinsic physical states of the subject—if the subject were in a radically different environment, the content ascribed to her visual states would be different.

The above argument applies to perceptual content in general. The cognitive tasks that define the domains of theories of perception are *typically* specified in terms of the recovery of certain types of information about the subject's normal environment. Interpreting the subject's perceptual states as carrying information about the environment will demonstrate that the theory has indeed answered the question it was initially set. Consequently, we should expect that the content ascribed to representational structures in the explanatory models of perceptual theories will be wide.

## 5. Concluding remarks

To summarize what I have argued: Computational theories are individualistic—the mental states characterized by computational theories are shared by *doppelgängers*. Representational contents play a role in computational psychology analogous to the role played by models in the physical sciences. The contents ascribed to mental states in explanatory models of computational theories of perception are, typically, wide; hence, such theories are *both* individualistic and externalist.

The account of content sketched above contrasts sharply with an alternative view, according to which semantic properties of representations play an explanatory role in a system's capacity to compute a cognitive function. According to the alternative proposal, when a system produces an early representation *R1*, as part of a process that culminates in the production of a later representation *R2*, *R1's having the content it does explains the production of R2*.<sup>19</sup> The alternative view, in my opinion, misplaces the explanatory contribution of content in computational accounts of cognitive capacities. Computational processes are blind to the semantic properties of the structures over which they are defined. A computational explanation of *R2's* production will appeal only to formal (i.e. non-semantic) properties of the system. Such explanations are *methodologically solipsistic*. But the fact that the system produces structures that are appropriately interpretable explains how a system that computes the hypothesized function could subserve the cognitive task that it does, where the task is typically described intentionally. The ascription of content plays an explanatory role in computational accounts, although not the role envisioned in the alternative view.

A final point: As previously noted, computational psychology respects the principle of modular design, treating cognitive processes as independently characterizable components of larger systems. The conspicuous successes of the discipline have been in the study of capacities that are relatively isolable, for example, early vision and syntactic and morphological analysis. Complex behaviour will be explained, if at all, as the interaction of multiple modular processes. The ascription of content serves an additional purpose besides its role in explanatory models of computational theories: it is essential for understanding how a module might be integrated into larger cognitive systems. Consider once again the *visex/audex* example. Recall that the two modules—one subserving visual perception, the other subserving auditory perception—are computationally identical. Characterizing what the components do in terms appropriate to the psychological domain in which they are normally deployed (that is, as either computing a representation of surface features of objects, or computing a representation of certain sonic properties) allows us to understand each module's role

<sup>19</sup> The content of the representations might be conceived as either wide or narrow. In some versions of the proposal, the content of early representations is explicitly claimed to be *causally efficacious* in the production of later representations



in the overall cognitive economy of the organism. Interpreting the inputs and outputs of modular processes seems unavoidable if we hope to explain how they interact to produce complex behaviour. Thus the choice of an explanatory model for a computational theory of a cognitive capacity is likely to be made with an eye to broader explanatory purposes—explaining how the output of a particular cognitive module feeds into later processes, and eventually explaining how the capacity contributes to the organism's successful interaction with its environment.<sup>20</sup>

Department of Philosophy  
Rutgers University  
New Brunswick  
NJ 08903  
USA

FRANCES EGAN

### REFERENCES

- Block, N. 1986: "Advertisement for a Semantics for Psychology", in *Midwest Studies in Philosophy Volume 10: Studies in the Philosophy of Mind*, French, P. A., Uehling, T. E. and Wettstein, H. K., eds., Minneapolis: University of Minnesota Press, pp. 615-78.
- Burge, T. 1979: "Individualism and the Mental", in *Midwest Studies in Philosophy Volume 4: Studies in Metaphysics*, French, P. A., Uehling, T. E. and Wettstein, H. K., eds., Minneapolis: University of Minnesota Press, pp. 73-121.
- 1986: "Individualism and Psychology". *The Philosophical Review*, 95, pp. 3-45.
- 1988: "Cartesian Error and the Objectivity of Perception", in *Contents of Thought*, Grimm, R. H. and Merrill, D. D., eds., Tucson: University of Arizona Press, pp. 62-76.
- Churchland, P. 1981: "Eliminative Materialism and the Propositional Attitudes". *Journal of Philosophy*, 78, pp. 67-89.
- Clark, A. 1989: *Microcognition*, Cambridge, Mass: MIT Press.
- Cummins, R. 1989: *Meaning and Mental Representation*, Cambridge, Mass: MIT Press.
- Davies, M. 1991: "Individualism and Perceptual Content". *Mind*, 100, 4, pp. 461-84.
- Dennett, D. C. 1987: "Evolution, Error, and Intentionality", in his *The Intentional Stance*. Cambridge, Mass: MIT Press, pp. 287-321.
- Egan, F. 1991: "Must Psychology Be Individualistic?" *The Philosophical Review*, 100, pp. 179-203.
- Fodor, J. A. 1987: *Psychosemantics*, Cambridge, Mass: MIT Press.

<sup>20</sup> Thanks to Robert Matthews, Colin McGinn, Gabriel Segal and Steve Stich for helpful comments on earlier drafts of this paper.

- Jackson, F. and Pettit, P. 1988: "Functionalism and Broad Content". *Mind*, 97, pp. 381-400.
- Kitcher, P. 1988: "Marr's Computational Theory of Vision". *Philosophy of Science*, 55, pp. 1-24.
- Kobes, B. 1990: "Individualism and Artificial Intelligence", in *Philosophical Perspectives, Vol.4: Action Theory and Philosophy of Mind*, J. Tomberlin, ed., Atascadero, CA: Ridgeview Publishing Co., pp. 429-459.
- Marr, D. 1981: "Artificial Intelligence: A Personal View", in *Mind Design*, Haugeland, J., ed., Cambridge, Mass: MIT Press, pp. 129-142.
- 1982: *Vision*. New York: Freeman Press.
- Marr, D. and Hildreth, E. 1980: "Theory of Edge Detection", in *Proceedings of the Royal Society*, London, B200, pp. 187-217.
- Matthews, R. 1990: "The Measure of Mind," report no. 57, *Research Group on Mind and Brain*, ZiF, Bielefeld.
- Segal, G. 1989: "On Seeing What is Not There". *The Philosophical Review*, 98, pp. 189-214.
- 1991: "Defence of a Reasonable Individualism". *Mind*, 100, 4, pp. 485-93.
- Stich, S. P. 1983: *From Folk Psychology to Cognitive Science*. Cambridge, Mass: MIT Press.
- Ullman, S. 1979: *The Interpretation of Visual Motion*. Cambridge, Mass: MIT Press.
- van Gulick, R. 1989: "Metaphysical Arguments for Internalism, and Why They Don't Work", in *Re-Representation*, Silvers, S., ed., Dordrecht: Kluwer Academic Publishers, pp. 151-9.