DUKE
UNIVERSITY PRESS

Philosophical Review

---

Computation and Content
Author(s): Frances Egan
Source: *The Philosophical Review*, Vol. 104, No. 2 (Apr., 1995), pp. 181-203
Published by: Duke University Press on behalf of Philosophical Review
Stable URL: http://www.jstor.org/stable/2185977
Accessed: 17/02/2011 14:54

---

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=duke.

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.

*Duke University Press* and *Philosophical Review* are collaborating with JSTOR to digitize, preserve and extend access to *The Philosophical Review.*

STOR®
http://www.jstor.org

*The Philosophical Review,* Vol. 104, No. 2 (April 1995)

# Computation and Content

## Frances Egan

## 1. Introduction

The dominant program in cognitive psychology since the demise of behaviorism in the 1960s has been *computationalism.* Computational theories treat human cognitive processes as a species of information processing, and the systems that implement such processing as symbol-manipulating systems. Describing a device as a symbol manipulator implies that it is possible to treat some of its internal states as representations of properties or objects in a particular domain. Computational theories of vision, for example, posit internal states that can be interpreted as representing the *depth* of the distal scene.

There has been considerable disagreement about the nature and function of representational contents assigned to the states posited by computational theories. It is widely thought that such theories respect what Jerry Fodor (1980) has called the "formality condition," which requires that computational processes have access only to the formal (that is, *nonsemantic*) properties of the representations over which they are defined. It is by respecting the formality condition that computationalism promises to answer one of the most pressing problems in the philosophy of mind—how can representational mental states be causally efficacious in the production of behavior? Representational mental states, according to computationalism, have their causal roles in virtue of (roughly) their structural properties.[1] But this advantage comes at a price. The formal character of computational description appears to leave no real work for the semantic properties of the mental states it characterizes. Thus, computationalism has been thought by some to support a form of *eliminativism,* the thesis that denies that intentionally characterized states play a genuinely explanatory role in psychology (see, for example, Stich 1983). If the content of computational states is indeed explanatorily idle, then the relation

---

[1] For language-like representations, the formality condition claims that they have their causal roles in virtue of their *syntax.*

between psychological states, as characterized by computational psychology, and psychological states as characterized by our commonsense explanatory practices, which do advert to content, is quite obscure.

In this paper I articulate and defend a strategy for reconciling the formal character of computational description with a commitment to the explanatory usefulness of mental content. I argue that content does not play an individuative or taxonomic role in computational theories—a computational characterization of a process is a *formal* characterization. Nonetheless, content does play a genuine explanatory role in computational accounts of cognitive capacities. Content ascriptions connect the formal characterization of an internal process with the subject's environment, enabling the computational theory to explain how the operation of the process constitutes the exercise of a cognitive capacity in that environment. I support my account of the role of content in computational psychology by reference to David Marr's theory of early vision,[2] in part because it has received a great deal of attention from philosophers; however, my argument depends on general features of computational methodology, and so applies to computational theories generally.

Recent attempts to reconcile computation and content have appealed to a notion of *narrow* content, that is, content that supervenes on intrinsic physical states of the subject. Proponents of narrow content have so far failed to articulate a notion that is clearly suitable for genuine explanatory work in psychology.[3] I argue that it is typically *broad* content that plays a central role in computational explanation, though I do identify a specific (and limited) function served by narrow content ascription.

## 2. Why Computational Theories Are Not Intentional

It might be argued that, the formality condition notwithstanding, computational theories are *intentional* in the following sense: The

---

[2] For the most detailed exposition of Marr's theory see Marr 1982.

[3] Loar (1988) and Segal (1989, 1991) have perhaps come closest. Loar's proposal concerns commonsense, as opposed to computational, psychology. Segal argues that narrow content plays a central role in Marr's theory. See Egan (forthcoming) for criticism of Segal's proposal.

states they posit not only have representational content, but the content they have plays an individuative role in the theory. In other words, computational theories taxonomize states by reference to their contents.

The motivation for the claim that computational theories of cognition are intentional in the above sense is not hard to understand. Consider the following passages:

> There is no other way to treat the visual system as solving the problem that the theory sees it as solving than by attributing intentional states that represent objective physical properties. (Burge 1986, 28–29)
>
> [I]t is at least arguable that where rational capacities are the *explananda,* it is necessary that there be propositional attitudes in the *explanans.* If this argument is correct, then it is pragmatically incoherent for Stich and his followers to insist that cognitive psychology explains rational capacities by reference to states not described as possessing propositional content. (Hannan 1993, 173)

The argument underlying both passages can be expressed somewhat crudely as follows:

(P) The *explananda* of computational psychological theories are intentionally characterized capacities of subjects.

(C) Therefore, computational psychological theories are intentional—they posit intentional states.

Underlying the argument is the intuition that scientific explanations should "match" (in some sense) their explananda. Wilson endorses a constraint of this sort, which he calls *theoretical appropriateness*:

> An explanation is theoretically appropriate when it provides a natural (e.g. non-disjunctive) account of a phenomenon at a level of explanation matching the level at which that phenomenon is characterized [in the explanandum]. (1994, 57)

The notion of a "level of explanation" is somewhat vague, but let us assume, for the sake of argument, that there is a unique level of explanation such that all and only explanations at that level involve ascriptions of content. If theoretical appropriateness is a desideratum of scientific explanation, then an explanation of intentionally characterized phenomena should itself advert to intentionally characterized states.[4]

An unresolved tension surfaces, though, when we consider computational explanation. The fact that the explananda of computational theories are intentionally specified suggests that computational states are essentially individuated by reference to their contents. If computational theories are not intentional, then how can computational theories explain intentionally characterized phenomena? But the formality condition exerts an opposite pressure. In requiring that computational processes have access only to the nonsemantic properties of the representational states over which they are defined, it suggests that computational individuation is nonsemantic, or in Fodor's terminology, *formal.* Are computational taxonomies intentional or formal? At this point it is helpful to turn to a well-developed example.

Interpreters of Marr's theory of vision have assumed that visual states are individuated in the theory by reference to their contents, hence that the theory is intentional (see Burge 1986; Kitcher 1988; Segal 1989, 1991; Davies 1991; Morton 1993; Shapiro 1993). Although there has been a good deal of disagreement about the sort of content (broad or narrow) that Marrian structures have, the assumption that content plays an individuative role in the theory has not been thought to require explicit argument.[5] Burge says that it is "sufficiently evident" that the theory is intentional from the fact that "the top levels of the theory are explicitly formulated in intentional terms" (1986, 35). I shall argue that in construing content as individuative, interpreters of Marr have misconstrued the role of content in computational theories.

While it is true that in his informal exposition of the various visual processes Marr typically characterizes them by reference to features of the distal scene, one should not read too much into

---

[4] A tacit appeal to theoretical appropriateness seems to underlie the argument of Graves et al. (1973) for the claim that the explanation of the speaker's knowledge of her language must appeal to internalized *knowledge* of grammar.

[5] Shapiro (1993) has described my claim (in Egan 1991) that Marr's theory is not intentional as "startling." It should not be startling. A central claim of Field (1978) is, as Field puts it in his (1986) paper, "that psychological theories have a non-intentional core" (114). In any event, interpreters of Marr have not defended the crucial assumption that his theory *is* intentional.

this fact. The processes are also characterized formally. They have to be—Marr's theory of vision is a computational theory, and a formal characterization guarantees that they are programmable (hence, physically realizable). The question is, which characterization does the individuative work?

Marr argued persuasively that an information-processing system should be analyzed at three distinct levels of description. The "top" level, which Marr called the *theory of the computation,* is a characterization of the function computed by the system—what the system does. The *algorithmic level* specifies an algorithm for computing the function, and the *implementation level* describes how the process is realized physically.[6] The top level in Marr's hierarchy is sometimes identified with Pylyshyn's semantic level (Pylyshyn 1984) and Newell's knowledge level (Newell 1982). In other words, the theory of the computation has been construed as essentially an intentional or semantic characterization of a mechanism. But such a construal makes somewhat puzzling Marr's insistence that the search for the algorithm must await the precise specification of the theory of the computation. He says, "unless the computational theory of the process is correctly formulated, the algorithm will almost certainly be wrong" (1982, 124), suggesting that the top level should be understood to provide a *function-theoretic* characterization of the device. Indeed, Marr explicitly points out that the theory of the computation is a mathematical characterization of the function(s) computed by the various processing modules. In describing the mathematical formula that characterizes the initial filtering of the image (the calculation of the Laplacian of the image convolved with a Gaussian), 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 characterization 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 $\nabla^2 G$ as an adequate descrip-

---

[6] In describing the levels of Marr's hierarchy as levels of *description* I do not mean to preclude treating the items classified by level as phenomena and processes rather than purely linguistic devices available to theorists.

> tion, although I take an unashamedly information-processing point of view. (1982, 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 of edge detection to detect *zero-crossings.* (A zero-crossing is a point where the value of a function changes its sign. Zero-crossings correspond to sharp intensity changes in the image.) Marr grants that the mathematical 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." More precisely, it is the description upon which the correct specification of the algorithm crucially depends.

The claim that the top level provides a mathematical characterization does not imply that it is wrong to speak of the visual system as taking representations of light intensity values as input and yielding representations of shape as output. I am not denying that computational processes have true intentional (semantic) descriptions. For some purposes, as we shall see in the next section, an intentional description of a process will be preferable to a formal characterization. It is not incorrect to say that an intentional characterization of the function computed by a mechanism resides at the top level in Marr's hierarchy, although the intentional characterization provides an *extrinsic* description of what the device does, and does not individuate the computational process. For the purpose of individuation, the precise mathematical description given by the theory of the computation is the description that counts.[7,8]

---

[7] In arguing for (various) intentional characterizations of the theory of the computation, interpreters of Marr point out that he speaks of the primal sketch as "representing the image," and of other structures as representing such distal properties as depth and surface reflectance. The assumption underlying such arguments is that Marr's words in these passages are decisive for settling issues of *taxonomy.* If theory interpretation were so simple, much of the philosophy of science would be out of business. The individuative principles of a scientific theory can rarely be read off the language used to articulate the theory. Marr is not generally careful or consistent in his language. There is no reason why he should be—he is not focusing on the issues that have concerned philosophers. In the passage I have quoted in the text, however, Marr is explicitly discussing fun-

If, as I have argued, the top level of a computational account provides a *purely mathematical* characterization of a device, then there is little temptation to construe the second, or *algorithmic* level, as intentional. (Only Burge, as far as I know, construes the algorithmic level as intentional, apparently because in discussing various possible algorithms Marr sometimes employs intentional language.) The algorithmic level of theory simply specifies *how* the function characterized in mathematical terms at the top level is computed by the system.

## 3. The Explanatory Role of Content

I have argued that Marr's theory of vision is not intentional. My argument appeals to general features of computational methodology; if I am right, then computational theories of cognition are not intentional—the states and processes characterized by such theories are not individuated by reference to the representational contents ascribed to them. The formal—namely mathematical—characterization does the taxonomic work.

Let us consider for a moment the implications of the claim that computational theories are not intentional. Two mechanisms that compute the same mathematical function, using the same algorithm, are, from a computational point of view, the same mechanism, even though they may be deployed in quite different environments. A computational description is an environment-independent characterization of a mechanism.[9] Inasmuch as compu-

---

damental commitments of the information-processing approach, in particular, how the theory of the computation is to be understood; so the passage bears special significance in the context of the current issue.

[8] Colin McGinn has pointed out to me that the theory of the computation is intentional in the following sense: it does specify an intended interpretation of a computational process—the intended interpretation is *mathematical.* The topmost level of a computational theory characterizes the system as computing a series of functions defined on mathematical entities. I am quite happy to say that a computational theory is intentional in this rather unusual sense. This is *not* the sense in which interpreters of Marr have taken his theory to be intentional. (They have assumed that the theory characterizes the system, essentially, as computing a function defined on aspects of the visual domain, and this is precisely what I deny.)

[9] This is not to suggest that the theorist can ignore the subject's environment in attempting to formulate a computational description of the device. Quite the contrary. See section 5.

tational processes are generally construed as *modular* processes, even the *internal* environment is irrelevant to the type-individuation of a computational process. Imagine a component of the visual system, called the *visex,* that computes a representation of the depth of the visual scene from information about binocular disparity.[10] 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 the theory of auditory processing appropriate to this creature, the audex computes a representation 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, they compute the same class of functions. The switch will make no discernible difference to the behavior of the creatures, nor to what is going on inside their heads. The two mechanisms are computationally identical, despite the difference in their normal internal environments.

It will perhaps be noted that the visual theory that describes the visex characterizes it as *computing a representation of depth from disparity,* and not as computing a representation of certain sonic properties, although it would do the latter if it were embedded in a different internal environment. The important point is that the postulated structures have no content considered independently of the environment (internal and external) in which they are normally situated. This is the sense in which an intentional characterization of a computational process is an *extrinsic* description. Structures in the raw primal sketch, which contains information from several distinct $\nabla^2 G$ channels and provides the input to most of the modular processes characterized by Marr's theory, are reliably correlated with such salient distal properties as object boundaries or changes in illumination, and are described by Marr as *representing* these properties. In some radically different environment, however, the same structures may be correlated with different distal properties, or perhaps with no objective feature of the world. In the latter world, the structures would not represent anything, except perhaps features of the *image.* They would have no distal content in that world.

---

[10] This is an adaption of an example from Davies 1991.

The point I wish to underscore is that an intentional characterization of a computational mechanism involves an implicit relativization to the context in which the mechanism is normally embedded. The mathematical characterization provided by the theory of the computation does not. Only the mathematical characterization picks out an essential property of a computational mechanism. The intentional characterization is not essential, since in some possible circumstances it would not apply.

What, then, is the role that representational content plays in computational accounts of cognitive processes, if not to essentially characterize cognitive processes? I have argued elsewhere (Egan 1992) that semantic interpretations play a role in computational psychology analogous to the role played by explanatory models in the physical sciences. There are two senses in which this is true. In the first place, an intentional characterization of an essentially formal process serves an expository function, explicating the formal account, which might not itself be perspicuous. Secondly, when a theory is incompletely specified (as is Marr's theory), the study of a model of the theory can often aid in the subsequent elaboration of the theory itself. A computational theorist may resort to characterizing a computation partly by reference to features of some represented domain, hoping to supply the formal details (i.e., the theory) later.

Though the analogy with models in physics is, I think, interesting and useful, the most important function served by intentional interpretations of computational processes is unique to psychology. The questions that antecedently define a psychological theory's domain are usually couched in intentional terms. For example, we want a theory of vision to tell us, among other things, how the visual system can detect *depth* from information contained in two-dimensional images. An intentional specification of the postulated computational processes demonstrates that these questions are indeed answered by the theory. It is only under an interpretation of some of the states of the system as representations of distal properties (like depth, or surface reflectance) that the processes given a mathematical characterization by a computational theory are revealed as *vision*. Thus content ascriptions play a crucial explanatory role: we need them to explain how the operation of a formally characterized process constitutes the exercise of a cognitive capacity in the environment in which the process is normally deployed.

Let us return for a moment to the argument considered earlier for the claim that computational theories of cognition are intentional:

(P) The *explananda* of computational psychological theories are intentionally characterized capacities of subjects.
(C) Therefore, computational psychological theories are intentional—they posit intentional states.

The premise of the argument is true—the questions that define a psychological theory's explanatory domain are usually couched in intentional terms—but, as we have seen, it does not follow that the theory characterizes the states and processes it describes as necessarily intentional. Computational states and processes will typically have no true intentional description when considered independently of an environment. Intentional characterizatons are therefore not part of the individuative apparatus of computational theories. In *this* sense, (C) is false. Yet the argument does contain an important insight: an intentional characterization is needed to connect a computational theory with its pretheoretic explananda. An explanation of how the visual system detects the depth of the scene from information contained in two-dimensional images is forthcoming only when the states characterized in formal terms by the theory are construed as *representations of distal properties.*

But, one might object, isn't this crucial explanatory role played by an intentional interpretation of a computational process enough to make the computational theory intentional? Indeed, it might seem that a computational theory, when divorced from the intentional interpretation that secures its explanatory relevance, cannot properly be characterized as a theory of *cognition.* There is a sense in which this is true; however, it does not undermine my point that computational theories are not intentional. Let me explain.

A computational theory provides a mathematical characterization of the function computed by a mechanism, but only in some environments can this function be characterized as a *cognitive* function (that is, a function whose arguments and values are epistemically related, such that the outputs of the computation can be seen as rational or cogent given the inputs). An example will make the point clearer. The matching of stereo images essential to the computation of depth from binocular disparity is aided, according to Marr, by a fundamental fact about our world—that disparity varies smoothly, because matter is cohesive. This is an example of what Marr calls a *natural constraint* (in particular, the *continuity* constraint). In some environments, the constraints that enable a cognitive interpretation of the mathematical function computed by a mechanism will not be satisfied. In environments where the continuity constraint is not satisfied (a *spiky* universe), the stereopsis module would compute the same formally characterized function, *but it would not be computing depth from disparity.* The function might have no cognitive (i.e., rational) description in this environment. A computational theory prescinds from the actual environment because it aims to provide an abstract, and hence completely general, description of a mechanism that affords a basis for predicting and explaining its behavior in any environment, even in environments where what the device is doing cannot comfortably be described as *cognition.* When the computational characterization is accompanied by an appropriate intentional interpretation, we can see how a mechanism that computes a particular mathematical function can, in a particular context, subserve a cognitive function such as vision.

A computational theory explains a cognitive capacity by subsuming the mechanism that has that capacity under an abstract computational description. Explaining a pretheoretically identifiable capacity by reference to a class of devices that have an independent, theoretical, characterization is an explanatory strategy familiar from other domains, particularly biology. The ability of sand sharks to detect prey is explained by positing within the shark the existence of an *electric field detector,* a device whose architecture and behavior is characterized by electromagnetic theory. Electromagnetic theory does most of the explanatory work in the biological explanation of the shark's prey detecting capacity. Of course, the explanation appeals to other facts—for example, that animals, but not rocks and other inanimate objects in the shark's natural environment, produce significant electric fields—but no one would suggest that such facts are part of *electromagnetic theory.* Similarly, by specifying the class of computational devices to which a mechanism belongs and providing an independent (i.e., noncognitive) characterization of the behavior of this class, a computational theory bears the primary explanatory burden in the explanation of a cognitive capacity. The intentional interpretation of the process also plays an explanatory role—it demonstrates that the capacity has been explained—but playing an essential role in the cognitive explanation does not thereby make it part of the *computational theory* proper.

So does it follow that computational theories are not cognitive? It depends. If a theory must give a cognitive characterization of a mechanism (according to which computing a cognitive function is a *necessary* property of the mechanism) to be a cognitive theory, then computational theories are not cognitive. If bearing the primary explanatory burden in an explanation of a cognitive capacity is sufficient, then they typically are.

Let us return briefly to Wilson's theoretical appropriateness condition, the requirement that a scientific explanation characterize a phenomenon at the same level in the explanans as in the explanadum. The above account of computational explanation suggests that theoretical appropriateness is not a general constraint on scientific explanation. Computational explanations characterize cognitive capacities in nonintentional, formal, terms. The requirement is independently implausible in any case, since it would rule out not only reductive explanations (e.g., microreductions) of antecedently characterized phenomena, but also explanation by functional analysis, the predominant form of explanation in both cognitive psychology and biology.[11] Such explanations typically analyze complex capacities or processes into more basic, less specialized, elements. For example, the explanation of the capacity to do long division appeals to the ability to copy numerals and perform multiplication and subtraction. The explanation of digestion appeals to more basic chemical processes. Both of these explanations appear to violate Wilson's theoretical appropriateness condition. Though Wilson grants that theoretical appropriateness is a *defeasible* constraint on scientific explanation, the ubiquity of explanations of this sort suggests that it is not a constraint at all.

## 4. The Ascription of Content

An interpretation of a computational system is given by an *interpretation function* $f_I$ that specifies a mapping between the postulated

---

[11] See Cummins 1983, chap. 2, for an account of functional analysis.

structures of the system and elements of some represented domain. For example, to interpret a device as an adder involves specifying an interpretation function $f_I$ that pairs states of the device with numbers. The device can plausibly be said to *represent* elements in the domain only if there exists an interpretation function that maps formally characterized structures to these elements in a fairly *direct* way.

Since an interpretation is just a structure-preserving mapping between formally characterized elements and elements of some represented domain, there is no reason to think that the interpretation of a computational system will be unique. The non-uniqueness of computational interpretation has been thought to be a problem for computationalism, but in fact it is not. Most "unintended" interpretations will not meet the directness requirement.[12] More importantly, the plausibility of a computational account depends only on the existence of an interpretation that does explanatory work.[13]

If the above account of the explanatory role of content is correct, then the interpretation of a computational system should connect the formal apparatus of the theory with its pretheoretic explananda. This requirement will constrain the choice of an appropriate interpretation function. A computational theory that purports to explain our arithmetical abilities cannot plausibly claim to have done so unless some of the states it postulates are interpretable as representing numbers. The fact that the system could also be interpreted as charting the progress of the Six-Day War (to use an example of Georges Rey's) would not undermine the theorist's claim to have described an arithmetical system, assuming that the mechanism can be consistently and directly interpreted as com-

---

[12] The directness requirement precludes interpreting a desk as an adder, since the assignment of numbers to states of the desk requires the interpreter to compute the addition function herself. The system is not doing the work. The directness requirement has yet to be precisely specified, but see Cummins 1989, chap. 8 for discussion. I gloss over this issue here primarily because, as we shall see below, the "problem" of ruling out unintended interpretations of computational systems typically does not arise.

[13] The existence of more than one interpretation meeting the directness requirement simply shows that the formally characterized device is capable of computing more than one cognitive function. The visex, described above, would compute a function on the auditory domain if it were embedded differently in the organism.

puting the appropriate arithmetical functions. Given the explanatory role of intentional interpretation as characterized in the previous section, the existence of "unintended" interpretations of computational systems is irrelevant. The preexisting explananda of the theory set the terms for the ascription of content.

Consider what this means for theories that purport to explain our perceptual capacities. 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 states of the system as representing environment-specific properties demonstrates that the theory explains how the subject is able to recover this information in its normal environment. Consequently, we should expect the contents ascribed to computationally characterized perceptual states to be *broad,* that is, not shared by physically identical subjects in significantly different environments.

It has been argued by Fodor (e.g., 1980, 1984, 1987) and others (e.g., Block (1986) and Cummins (1989)) that computational psychology must restrict itself to a notion of *narrow* content, that is, content that supervenes on intrinsic physical states of the subject.[14] In part, the motivation for such a view is the recognition that computational taxonomy prescinds from the subject's normal environment. Physical duplicates are computational duplicates. Given this fact, if computational states have their semantic properties essentially, then computational psychology requires a notion of content that supervenes on the physical properties of the system; in other words, it needs a notion of narrow content. But if, as I have argued, computational states have their semantic properties only *nonessentially,* then narrow content is not necessary. And it turns out that there are good reasons why computational psychology should not restrict itself to narrow content.

In the first place, a useful notion of narrow content has been notoriously hard to specify. More importantly, since the explananda of theories of perception are typically formulated in environment-specific terms, ordinary environment-specific broad contents will best serve the explanatory goals of such theories. The point

---

[14] Others, such as Stich (1983), impressed by the fact that content ascription is typically context-sensitive and observer-relative, have concluded that cognitive psychology should not advert to content at all.

can be generalized. It is widely appreciated that ordinary contents are broad. Insofar as the pretheoretic explananda of computational theories are framed in ordinary terms, the ascription of broad content to computational states and structures will be appropriate.

A close look at Marr's theory confirms the point. He ascribes broad, environment-specific contents where possible. If in a subject's normal environment a structure is reliably correlated with a salient distal property, then Marr describes the structure as representing that property. (For example, he describes structures in the 2.5D. sketch as representing *surface orientation.*) Some of the structures posited by Marr's theory correlate with no simple distal property tokening in the subject's normal environment. The structures that Marr calls *edges* sometimes correlate with changes in surface orientation, sometimes with changes in depth, illumination, or reflectance. Marr describes edges as representing this disjunctive distal property. Notice that in both cases—correlation with a simple distal property in the subject's normal environment or correlation with a disjunctive distal property in the subject's normal environment—the contents ascribed to the representational structures are broad. Moreover, the broad contents so ascribed are determined by the correlations that obtain in the subject's normal environment, not by those that would obtain in some other environment.

Some of the structures that Marr posits (e.g., individual zero-crossings) do not, however, correlate with any easily characterized distal property, simple or disjunctive, in the subject's normal environment. Some of their tokenings correlate with distal properties, others appear to be mere artifacts of the imaging process. Marr recognizes this fact, cautioning that such structures as zero-crossings are not "physically meaningful"; he describes them as representing *discontinuities in the image.* Their contents are only proximal, and hence narrow—they supervene on the intrinsic properties of the subject. But such proximal or narrow content, far from being Marr's content of choice, is his content of last resort, since he ascribes proximal content only when a broad content ascription is unavailable.[15]

---

[15] Commentators who have thought narrow content to be Marr's content of choice have presumably done so because they recognize that content-determining correlations with distal properties can vary wildly across environments (see, for example, Segal 1989, 1991). They fail to notice that

*Covariational* (or *information-theoretic*) theories of content identify the meaning of a representational state with the cause of the state's tokening in certain specifiable circumstances.[16] The foregoing account of content ascription in Marr's theory may tempt some to find in his theory a tacit endorsement of a covariational theory of content. This would be a mistake. I have claimed that in ascribing content Marr looks for salient distal correlates of a structure's tokening in the subject's normal environment. I have been careful to avoid claiming that these correlates are *the cause* of the structure's tokening. Though it may be natural to say that they are, Marr makes no such claim, and a number of well-known problems are avoided by not doing so.[17] It should be clear that Marr's theory is not committed to a covariational theory of content if one considers the sort of case where no *salient* distal correlate (simple or disjunctive) of a structure's tokening can be found. In such cases,

---

for Marr the relevant correlations are those that obtain in the subject's normal environment.

[16] See, for example, Stampe (1977), Dretske (1981), and Fodor (1990). There are, of course, important differences in their accounts.

[17] One problem with covariational theories of content is their implication that the meaning of a symbol is given by the disjunction of all of its potential causes. Since "horse" tokenings would be caused not only by horses, but also by horsey looking cows, covariational theories seem to imply that "horse" means *horse or horsey looking cow.* (See Fodor 1990 for discussion.) The "disjunction problem" gives rise to a further difficulty, namely, how to account for the possibility of a symbol's *mis*representing its object, given that all potential causes of a symbol's tokening determine its meaning. Though computational theorists have had little to say about misrepresentation (their concern is to characterize what is going on in the normal case, where perception is veridical) it is not hard to see how misrepresentation can arise on the account of content ascription I have sketched above. Structures assigned distal contents (simple or disjunctive) will misrepresent if they are tokened when the normal environmental conditions for their tokening are not satisfied. Suppose that, as part of a military training exercise, Bill, a normal human with a Marrian visual system, is placed in a room where the continuity constraint, which holds that disparity varies smoothly because matter is cohesive, is not satisfied. Bill's visual system normally computes *depth* from disparity information. However, in these circumstances, where spikes of matter project in all directions, Bill (or, more specifically, the stereopsis module of Bill's visual system) will compute the same formally characterized function as he normally does, but he will misrepresent some other property (not a property for which we have a convenient name) as depth. In general, where the contraints that normally enable an organism to compute a cognitive function are not satisfied, it will fail to represent its environment.

Marr ascribes a proximal content to the structure, interpreting it as representing a feature of the image or input representation rather than the distal cause of its tokening, whatever that might be. The ascription of proximal content serves an important expository purpose—it makes the computational account of the device more perspicuous, by allowing us to keep track of what the device is doing at points in the processing where the theory posits structures that do not correlate neatly with a salient distal property.[18] No explanatory purpose would be served by an unperspicuous distal interpretation of these structures; consequently, Marr does not interpret them as representing their distal causes. The decision to adopt a proximal rather than a distal interpretation is dictated by purely explanatory considerations.[19]

## 5. Computational Psychology and Naturalistic Psychology

The compatibility of computational description and broad content seems to have gone unnoticed in the literature. Cummins (1989), whose account of representation in computational psychology bears some resemblance to mine, says the following:

> The CTC [computational theory of cognition] . . . seeks an *individu-*

---

[18] One might wonder whether what I am calling "proximal content" is really content at all. To be sure, proximal contents do not bear much resemblance to the contents we ascribe in our ordinary predictive and explanatory practices; however, I do think that contents of this sort play a genuine explanatory role in computational accounts of internal processes. To cite a second example, Marcus (1980) interprets the structural descriptions constructed in the course of natural language comprehension as representing not distal objects (or public language sentences) but the items in stacks or buffers of the parser. In both the vision and parsing cases, interpreting a structure as representing other structures constructed earlier in the process serves the important function of allowing us to keep track of what the processor is doing. Given that the rationale for content ascription in computational psychology is primarily explanatory, I think that proximal content should be treated as a species of content, though perhaps only as a sort of "minimal" content.

[19] Matthews (1988), Segal (1989), and McGinn (1989) note another reason to resist an exclusively causal account of content. They argue that the contents of mental representations seem to be partly determined by the sorts of behaviors that they tend to produce. Whether a structure whose tokening is caused by both cracks and shadows means *crack*, *shadow*, or *crack or shadow* depends in part upon whether its tokening contributes to the production of behavior appropriate to cracks, shadows, or both.

> *alist* psychology, i.e., a psychology that focuses on cognitive capacities of the kind that might be brought to bear on radically different environments. If the anti-individualist position with regard to intentionality is right (i.e. if beliefs and desires cannot be specified in a way that is independent of environment), then the explananda of an individualist psychology cannot be specified intentionally. It follows that the CTC shouldn't—indeed, *musn't*—concern itself with intentionally specified explananda. (140)

Cummins's mistake is in thinking that the fact that a computational theory seeks to provide a nonintentional, environment-independent characterization of a cognitive process entails that it cannot explain phenomena specified in environment-specific terms. This, we have seen, is wrong. A computational theory explains an environment-specific cognitive capacity by subsuming it under an environment-independent characterization. The intentional interpretation of the process serves as a bridge between the abstract characterization provided by the theory and the environment-specific intentional characterization that constitutes the theory's explananda. Precisely because the intentional interpretation does not play an essentially individuative role in the theory—in other words, whatever contents computational states have, they have them *nonessentially*—the theorist is free to assign broad contents where appropriate to secure the connection between theory and explananda.

The fact that the computational theorist can and typically will assign broad contents to computational structures has larger implications for psychology. In "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology," Fodor says the following:

> there is room for both a computational psychology—viewed as a theory of formal processes defined over mental representations—and a naturalistic psychology, viewed as a theory of the (presumably causal) relations between representations and the world which fix the semantic interpretations of the former. I think that in principle this is the right way to look at things . . . however . . . it's overwhelmingly likely that computational psychology is the only one that we are likely to get. . . . [A] naturalistic psychology isn't a practical possibility and isn't likely to become one. (1980, 66)

Naturalistic psychology, as Fodor construes it, is the theory of organism/environment relations that fix the meanings of our mental terms. He offers two arguments for the claim that a naturalistic psychology is impossible. As both arguments have been thoroughly worked over in the literature (see the commentaries that accompany Fodor 1980), I won't go into them here. But as far as I know, no one has disputed Fodor's implication that computational psychology and naturalistic psychology are entirely unrelated projects. If my account of the role of content in computational psychology is correct, then Fodor's way of conceiving things is wrong. If we had a complete computational psychology, that is, a computational account of each human cognitive capacity, we would *ipso facto* already have a naturalistic psychology. Let me elaborate.

Although a computational theory provides a formal, environment-independent, characterization of a process, the theorist will usually be unable to discover the correct formal characterization without investigating the subject's normal environment. Typically, a necessary first step in specifying the function computed by a cognitive mechanism is discovering environmental constraints that make the computation tractable. The solutions to information-processing problems are often underdetermined by information contained in the input to the mechanism; the solution is achieved only with the help of additional information reflecting very general features of the subject's normal environment. For example, as previously mentioned, the computation of depth from binocular disparity is possible only because the mechanism is built to assume something that is true about its normal environment—that disparity varies smoothly because matter is cohesive (the continuity constraint). Finding constraints of this very general sort is a necessary first step in characterizing the mathematical problem that the mechanism has to solve, and thus in arriving at a correct computational description of the process.

There is a second and more obvious point at which the computational theorist will contribute to the specification of the organism/environment interactions that fix the meanings of mental terms—namely, in the specification of an intentional interpretation of a formally characterized process. I have argued that content ascription is constrained by the subject's normal environment. The process of ascribing content to the structures posited by the theory involves the attempt to specify the normal environmental correlates of tokenings of these structures. The fact that Marr succeeded in ascribing distal contents to many of the structures posited in his theory (and Marr is not unique in this achievement) demonstrates that naturalistic psychology is not impossible. Although computational psychology is formal—its taxonomic principles are formal—it develops hand-in-glove with the project that Fodor calls naturalistic psychology.[20]

## 6. Scope and Limits of the Account

My account of the explanatory role of content has been articulated and defended by reference to *classical* computational models. Classical architectures treat cognitive processes as rule-governed manipulations of internal symbols or data structures that are explicit candidates for interpretation. *Connectionist* cognitive models, by contrast, do not posit data structures over which the device's operations are defined. (Many connectionist devices, unlike classical devices, are not correctly described as constructing, storing, and retrieving internal representations.) Connectionist models posit activated units (nodes) that increase or decrease the level of activation of other units to which they are connected until the ensemble settles into a stable configuration. Consequently, connectionist models lack convenient "hooks" on which an intentional interpretation of a process may be hung. Semantic interpretations ascribe content either to individual units in the network or to patterns of activation over an ensemble of units. However, I see no reason why the above account of the explanatory role of content would not apply straightforwardly to connectionist systems. Semantic interpretations of connectionist networks play the same complex explanatory role as do interpretations of classical computational models. Most importantly, they connect a connectionist theory of a cognitive capacity with its pretheoretic explananda.

It remains to be seen whether computational psychology will shed much light on paradigm cases of intentional states, namely, beliefs and desires. The conspicuous successes of computationalism

---

[20] Naturalistic psychology, construed as the specification of the organism/environment relations that fix the meanings of mental representations, should not be confused with the enterprise that is sometimes called "the naturalization project" in semantics. The latter attempts to specify sufficient conditions, in a nonintentional and nonsemantic vocabulary, for a mental state's meaning what it does. It is a purely philosophical project, not the concern of psychologists.

have been in characterizing highly modularized, informationally encapsulated processes such as early vision and syntactic and phonological processing. The states posited by theories of this sort fail to exhibit the complex functional roles characteristic of the propositional attitudes (including, typically, accessibility to consciousness). They are *subdoxastic* states. Fodor, in *The Modularity of Mind,* has expressed considerable pessimism about the prospects of characterizing in formal, computational terms more *central* cognitive processes such as belief fixation. I think this pessimism is well placed, if only because the context-sensitivity of belief ascription makes the programming task appear intractable. However, should a computational account of propositional attitudes be forthcoming, content would play the same explanatory role it plays in theories of modular capacities. An interesting consequence of this eventuality is that propositional attitudes, so characterized, would not have their contents essentially. Type-identical belief-state tokens might have different contents, should they be tokened in relevantly different environments.[21] The prospect of a computational theory of belief, therefore, challenges a fundamental commitment of orthodox philosophy of mind.[22] Some may conclude that such a theory would not really be about the propositional attitudes, though nothing in the *folk* conception of the mind would seem to warrant this conclusion.[23]

*Rutgers University*

---

[21] For example, a computational theory of belief would type-identify my *water* beliefs and my Twin Earth doppelganger's *twater* beliefs, although intentional interpretations appropriate to our respective worlds might assign different broad contents to our type-identical beliefs. A computational theory of belief would, therefore, respect the intuition that has been the prime motivation for the postulation of narrow content—that doppelgangers are identical in psychologically relevant respects, and hence should be subsumed under the same psychological generalizations. But because a computational theory is not committed to narrow content, it can also accommodate the intuition that the subject's environment is a determinant of her belief *contents*.

[22] But see Matthews 1994 for an account of propositional attitudes that denies that they have their contents essentially.

[23] Thanks to Kent Bach, Noam Chomsky, Patricia Kitcher, Robert Matthews, Colin McGinn, Rob Wilson, and the editors for helpful comments on earlier versions of this paper.

## References

Block, Ned. 1986. "Advertisement for a Semantics for Psychology." In *Midwest Studies in Philosophy,* vol. 10, ed. P. French, T. Uehling, and H. Wettstein. Minneapolis: University of Minnesota Press.

Burge, Tyler. 1986. "Individualism and Psychology." *Philosophical Review* 95:3–45.

Cummins, Robert. 1983. *The Nature of Psychological Explanation.* Cambridge: MIT Press.

———. 1989. *Meaning and Mental Representation.* Cambridge: MIT Press.

Davies, Martin. 1991. "Individualism and Perceptual Content." *Mind* 100: 461–84.

Dretske, Fred. 1981. *Knowledge and the Flow of Information.* Cambridge: MIT Press.

Egan, Frances. 1991. "Must Psychology be Individualistic?" *Philosophical Review* 100:179–203.

———. 1992. "Individualism, Computation, and Perceptual Content." *Mind* 101:443–59.

———. Forthcoming. "Intentionality and the Theory of Vision." In *Problems in Perception,* ed. K. Akins. Oxford: Oxford University Press.

Field, Hartry. 1978. "Mental Representation." *Erkenntnis,* 13:9–61.

———. 1986. "The Deflationary Conception of Truth." In *Fact, Science, and Morality: Essays on A.J. Ayer's "Language, Truth, and Logic,"* ed. G. McDonald and C. Wright, 55–117. Oxford: Basil Blackwell.

Fodor, Jerry. 1980. "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology." *Behavioral and Brain Sciences* 3:63–73.

———. 1984. "Observation Reconsidered." *Philosophy of Science* 51:23–43.

———. 1987. *Psychosemantics.* Cambridge: MIT Press.

———. 1990. *A Theory of Content and Other Essays.* Cambridge: MIT Press.

Graves, C., J. Katz, Y. Nishiyama, S. Soames, R. Stecker, and P. Tovey. 1973. "Tacit Knowledge." *Journal of Philosophy* 70:318–30.

Hannan, Barbara. 1993. "Don't Stop Believing: The Case Against Eliminative Materialism." *Mind and Language* 8:165–79.

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

Marcus, Mitchell. 1980. *A Theory of Syntactic Recognition for Natural Language.* Cambridge: MIT Press.

Marr, David. 1982. *Vision.* New York: Freeman Press.

Marr, David, and Ellen Hildreth. 1980. "Theory of Edge Detection." *Proceedings of the Royal Society* B200:187–217.

Matthews, Robert. 1988. "Authoritative Self-Knowledge and Perceptual Individualism." In *Contents of Thought,* ed. R. Grimm and D. Merrill. Tucson: University of Arizona Press.

———. 1994. "The Measure of Mind." *Mind* 103:131–46.

McGinn, Colin. 1989. *Mental Content.* Oxford: Basil Blackwell.

Morton, Peter. 1993. "Supervenience and Computational Explanation in Vision Theory." *Philosophy of Science* 60:86–99.

Newell, Allen. 1982. "The Knowledge Level." *Artificial Intelligence* 18:87–127.

Pylyshyn, Zenon. 1984. *Computation and Cognition.* Cambridge: MIT Press.

Segal, Gabriel. 1989. "On Seeing What is Not There." *Philosophical Review* 98:189–214.

———. 1991. "Defence of a Reasonable Individualism." *Mind* 100:485–93.

Shapiro, Lawrence. 1993. "Content, Kinds, and Individualism in Marr's Theory of Vision." *Philosophical Review* 102:489–513.

Stampe, Dennis. 1977. "Toward a Causal Theory of Linguistic Representation." In *Midwest Studies in Philosophy,* vol.2, ed. P. French, T. Uehling, H. Wettstein. Minneapolis: University of Minnesota Press.

Stich, Stephen. 1983. *From Folk Psychology to Cognitive Science.* Cambridge: MIT Press.

Wilson, Robert. 1994. "Causal Depth, Theoretical Appropriateness, and Individualism in Psychology." *Philosophy of Science* 61:55–75.