© *Blackwell Publishers Ltd. 1999, 108 Cowley Road, Oxford, OX4 1JF, UK*
and 350 Main Street, Malden, MA 02142, USA.
Mind & Language, ISSN: 0268-1064
Vol. 14. No. 2 June 1999, pp 177–194.

In Defence of Narrow Mindedness

FRANCES EGAN

Abstract: Externalism about the mind holds that the explanation of our representational capacities requires appeal to mental states that are individuated by reference to features of the environment. Externalists claim that 'narrow' taxonomies cannot account for important features of psychological explanation. I argue that this claim is false, and offer a general argument for preferring narrow taxonomies in psychology.

There are two opposing viewpoints concerning the individuation of psychological states. One, known as *individualism*, holds that the behaviour and rational capacities of agents are to be explained by reference to states that supervene on internal physical states of the agent; in other words, they are individuated *narrowly*. Its advocates include Fodor (1980, 1987), Block (1986), and Segal (1989, 1991). The denial of this view, known as *externalism* or *anti-individualism*, holds that explanatory states and constructs in psychology make essential reference to features of the subject's environment; they are *widely* individuated. According to externalism, physically identical subjects might be psychologically different. This position is championed by, among others, Burge (1986), Davies (1991), Millikan (1984), Papineau (1987, 1993), Peacocke (1994), Shapiro (1993, 1997), and Wilson (1994, 1995).

Externalism is clearly in the ascendancy. Individualism seems passé, a remnant of a stubborn Cartesianism that refuses to face the obvious fact that minds are embedded in a larger world (see Wilson, 1995). Can it really be denied that the character of mental processes depends on the environment in which they have developed and to which they are adapted? I shall argue that, in an important sense, this claim can and should be denied.

The dispute between externalists and individualists has tended to focus on the nature of mental *content*, since mental states (excepting, perhaps, men-

Earlier versions of this paper were presented at a symposium on Content and Computation, School of Advanced Study Philosophy Programme, University of London, and to the philosophy departments at MIT and the University of Rochester. Thanks to fellow London symposiasts Jose Bermudez and Sarah Patterson, and to audience members on these occasions, for helpful comments. Thanks also to an anonymous referee for this journal.

Address for correspondence: Frances Egan, Department of Philosophy, Rutgers University, Davison Hall, Douglass Campus, PO Box 270, New Brunswick, New Jersey, NJ 08901-0270, USA.
Email: fegan@rci.rutgers.edu.

tal states whose characteristic feature is their qualitative 'feel'[1]) are assumed by both sides to be individuated by their representational contents. If mental content is externalist—if it is individuated by reference to the subject's environment—then so are mental states that have externalist content. I have argued in a series of papers (Egan, 1991, 1992, 1995) that, at least with respect to *computational* theories of mind, the assumption that the individualism issue can be settled by focusing on the nature of representational content is false. Computational theories are *formal* in the following sense: the content ascribed to mental states by a computational cognitive theory—what these states *represent*—plays no role in the individuation of the states and processes postulated by the theory. Whatever the nature of mental content, computational theories are individualistic.

Externalists typically argue that unless psychological states are construed as widely individuated, psychology cannot perform the explanatory duties required of it. These duties include explaining the fact that an organism's perceptual and motor systems are adapted to its environment. The fact that a mechanism is adapted to a particular environment is, of course, a non-individualistic fact about it; it is a relational fact. A physically identical mechanism with a different history of selection pressures, or no such history, would lack this feature. Sometimes it is the fact that the tasks performed by cognitive mechanisms are specified in non-individualistic terms that is taken to be decisive. The overall task of the visual system, for example, is the specification of the shape and location of objects in the organism's environment (see Shapiro, 1993, 1997). In a similar vein, Peacocke (1994) argues that psychology is in the business of explaining not only behaviour, but also, perhaps even primarily, the intentional states of agents, and intentional states are distinguished by the fact that they have content. Contentful states cannot be explained by reference to states that are not themselves contentful, it is claimed, hence the thesis that computational individuation is formal, in the specified sense, must be wrong. Furthermore, since content itself is widely individuated, making essential reference to features of the subject's physical and social environment,[2] then the states that figure in psychological explanation must be widely (that is, externally) individuated.

All these arguments have in common the claim that psychological states must be widely individuated because the explananda of psychological theory are widely specified, or in some way essentially tied to the environment. Those who insist that psychological individuation is generally narrow are therefore thought to be overlooking important features of psychological explanation.

This is how I plan to proceed: I shall argue first that formally, or non-semantically, individuated computational states can and do play a central role

[1] Although see Dretske, 1995, and Tye, 1995, for representational accounts of qualitative mental states.

[2] This is the lesson of the Putnam/Burge thought experiments.

in explanations of the intentional states of agents; that is, states that have representational content. If my account of computational explanation is correct, then we can see that wide or externalist specification of psychology's explananda is consistent with the narrow individuation of its explanatory states. Then I offer a general argument for preferring narrow individuation over taxonomies that make essential reference to features of the environment.

1. An Apparent Inconsistency

On one view of computation, articulated by Fodor (1980), computational processes are sensitive only to the non-semantic properties of the representations over which they are defined. Such processes have no access to, for example, what a representation means, or what feature(s) of the environment it might be about. Peacocke (1994) argues that this 'non-semantic' view of computation involves a contradiction:

> If the non-semantic view of computation were correct, it certainly looks as if there would have to be a massive mismatch between means and ends in much contemporary psychology. It looks for all the world as if much theorizing in psychology attempts to explain particular intentional, content-involving properties of a subject . . .
>
> On the non-semantic view of computation, a computational explanation of a person's coming to be in an intentional state involves one non-semantic state explaining, by some computational procedure, a second non-semantic state. This second state is said to be the basis of (or realization of, or what constitutes) the intentional state to be explained. But if only non-semantic properties are explained, where is the explanation of the intentional properties? It seems that on the non-semantic conception of computation, only non-semantic features of intentional states could be explained. (p. 304)

Peacocke attempts to resolve the apparent inconsistency by proposing a different conception of computation—what he calls 'content-involving computation'—according to which the content of an internal state or event is computed from the contents of earlier states or events according to a 'content-involving algorithm'. I am not going to attack Peacocke's alternative conception of computation, save to say that the crucial notion of a 'content-involving algorithm' is not sufficiently spelled out. Rather, I want to undercut the motivation for an alternative account by challenging the claim that the non-semantic account of computation gives rise to an inconsistency.

2. Computational Explanation—A Non-Semantic Account

By virtually all accounts,[3] an interpretation of a computational system is given by an *interpretation function* f_I that specifies a mapping between equiv-

[3] See, for example, Fodor, 1975; Pylyshyn, 1984; and Cummins, 1989.

alence classes of physical states of the system and elements of some represented domain. For example, to interpret a device as an adder involves specifying an interpretation function that pairs states of the device with numbers. To interpret a device as a visual system requires specifying a mapping between states of the device and visible properties in the immediate environment. The device can plausibly be said to *represent* elements in the domain only if there exists an interpretation function that maps states of the device to these elements in a fairly direct way.[4]

A computational theory gives a formal characterization of a cognitive capacity. Computational states are individuated by a computational theory without essential reference to the contents assigned to them by the interpretation function. In other words, computational states do not have their contents essentially. Let me spell out the implications of this claim.

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 different environments. A computational description is an environment independent characterization of a mechanism. Thus, to take a familiar example, the states and structures characterized by computational vision theories are said to *represent* certain properties of the distal scene; structures in what David Marr called the 2.5D sketch represent depth and surface orientation (see Marr, 1982). However, computational vision theories individuate these states formally, and so independently of the environment in which the visual system is normally deployed, and to which it is adapted. The semantic interpretation of the mechanism, provided by the appropriate interpretation function, is an extrinsic description. If a Marrian visual system were somehow (say by random mutation) to appear in a radically different environment to which it was not adapted, then the same computationally described mechanism might not be correctly described by the semantic characterization that is appropriate to our world. Suppose that the states and structures posited by the theory do not covary with the same distal properties (changes in depth and surface orientation) in the counterfactual world. Perhaps they covary with different distal properties, or with some large disjunction of distal properties. Then tokenings of these structures would not represent depth and surface orientation in the counterfactual world. Given what I have said, we cannot say what they would represent in the counterfactual world; perhaps they would represent only features of the retinal image. But we do know that the device would still compute a well-defined mathematical function, specified by the

[4] The directness requirement needs to be precisely specified. It requires at a minimum that independently characterized states of the device covary with elements of the intended domain. The requirement that the mapping be direct precludes interpreting the wall behind me as an adding machine, since the assignment of numbers to states of the wall requires the interpreter to compute the addition function herself. The system is not doing the work.

computational theory. This description is true of the device independently of the environment in which it is deployed.[5]

It might be objected that the output of a computational mechanism depends on the mechanism's computing a particular *physical magnitude*, which it will do only in certain environments, hence the computational description is not environment independent. The device responsible for computing depth from retinal disparity, for example, would compute a different magnitude in an environment in which light rays converge over bent pathways. This is true, but it does not follow that the computational description is environment *dependent*, in the sense at issue. If the environment were different, this device would compute the same mathematical output, but it would not compute the same physical magnitude, that is, depth. The mathematical output computed by the device could not be interpreted as a specification of depth. The device that computes the mathematical function specified by the computational description would, in the environment in question, fail to compute depth from disparity. Of course, we might wonder about the adaptive value of such a device in this counterfactual environment, but cognitive mechanisms can be assumed to be adaptive only in the actual environment.

A crucial assumption of the computational approach, as I have described it, is that the fact that a mechanism is adapted to its environment is a *non-essential* property of it, *qua* computational mechanism. A computational theory, in providing a formal characterization of a device, abstracts away from the device's historical properties. The theorist attempting to characterize the cognitive capacities of adapted organisms must, of course, attend to the structure of the organism's environment to discover the computational problems that the organism, in its natural environment, needs to solve—otherwise a computational theory will be a nonstarter as a biological model—but the computational characterization is itself 'environment neutral'.[6]

Computation, then, is *non-semantic*. While computational states *have* content, their content is not essential to their identity as computational states. A computational process is, typically, a transition from one contentful state to another, but the process is literally a *computation*, a calculation; in no clear sense is computation 'content involving'.

But then how are computational theories, so construed, able to explain the intentional states of agents? According to Peacocke they cannot. And what

[5] The claim that a computational characterization is formal, or non-semantic, needs some clarification. Given that a computational characterization of a device specifies the mathematical function computed by the device, the computational characterization *is* a semantic characterization. But mathematical characterization is not what theorists typically have in mind when they talk about 'the semantic interpretation of a device'. The semantic interpretation specified by a computational theory of vision, for example, will assign *visual* contents to the states it characterizes, and the computational characterization prescinds from these contents.

[6] See Egan, 1995, for elaboration of this point.

is the role that representational content plays in computational accounts of cognitive processes, if not to essentially characterize cognitive processes?

Content ascription serves several important explanatory functions. I have suggested (in 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 a computational process serves an expository function, explicating the formal account which might not itself be perspicuous. Secondly, when a theory is incompletely specified (as is the case with 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. In the meantime, contents can serve a reference-fixing or indexing function, allowing the theorist to refer to states yet to be given a precise formal characterization.

I think that the analogy with models in physics is interesting and useful, but it doesn't help us to resolve Peacocke's worry—how a non-intentional theory could explain intentional states. The most important function served by a semantic interpretation of a computational process 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 three-dimensional distal structure from information contained in two-dimensional images. An intentional characterization of the postulated computational processes enables the theory to answer these questions. The semantic interpretation tells us that states of the system covary, in the normal environment, with changes in depth and surface orientation. It is only under an interpretation of some of the states of the system as representations of depth and surface orientation that the processes given a formal characterization by a computational theory are revealed as *vision*. Thus, content ascription plays a crucial *explanatory* role: it is necessary 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. The device would compute the same mathematical function in any environment, but only in some environments would its doing so enable the organism to *see*.

To summarize the point: an explanation of how the visual system detects the depth of the scene is forthcoming only when the states characterized in formal terms by the theory are construed as *representations of distal properties*. 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 explaining and predicting its behaviour in any environment, even in environments where we cannot say what, if anything, the device represents. When the computational characterization is accompanied by an appropriate semantic interpretation, we can see how a mechanism that computes a certain mathematical function can, in a particular context, subserve a cognitive function such as vision. We can talk of contents being 'computed' if we like, as long as we recognize that such talk is loose.

An oft-noted fact about the interpretation of computational mechanisms is that it is not unique, since an interpretation is just a structure-preserving mapping between formally characterized elements and elements of some represented domain.[7] (Of course, if computation is construed as *essentially* content-involving, as it is on Peacocke's account, then, trivially, a computational mechanism has a unique interpretation.) However, the non-uniqueness of semantic interpretation poses no problem for computational theories. The plausibility of a computational account of a cognitive capacity depends only on the existence of an interpretation that does genuine explanatory work. Let me elaborate.

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 pre-theoretic explananda. This requirement will constrain the choice of an appropriate interpretation. As noted above, a computational theory that purports to explain our visual abilities cannot plausibly claim to have done so unless some of the states it posits are interpretable as representing visible properties of the distal scene. This means that internal states given an independent characterization by the theory must covary with, for example, the depth and orientation of objects in the visual field. The computational states must *track* the appropriate distal properties. It is possible, though unlikely, that these states also covary with the fluctuating stock-market index, or plausible moves in a chess game, and hence that the system could be interpreted as keeping track of the stock market or playing a decent game of chess. But this possibility does not undermine the theorist's claim to have described a visual system, assuming that the system can be consistently and directly interpreted as computing the appropriate functions on the visual domain. Given the explanatory role of an intentional interpretation, the existence of 'unintended' interpretations is irrelevant. The pre-existing explananda of the theory determine the appropriate domain for the ascription of content. In the absence of a causal connection between the device and the stock market, or the opponent's chess moves, these 'accidental' correlations, or the theoretical possibility of such correlations, are of no interest.

The foregoing account has implications for the wide vs. narrow content dispute. It has been argued by Fodor (e.g. 1980, 1987) and others (e.g. Block, 1986; 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. In part, the motivation for such a view is the recognition that computational taxonomy prescinds from the subject's nor-

[7] This fact is often cited as an objection to computational models as accounts of 'original' or 'intrinsic' intentionality.

mal environment. Computational states supervene on the intrinsic physical states of the subject possessing them. Given this fact, if computational states have their semantic properties essentially, then computational psychology requires a notion of content that supervenes on intrinsic 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 *non-essentially*, 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.[8] More importantly, 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 wide (or externalist); that is, not necessarily shared by physically identical duplicates in different environments. Putting the point another way, since the explananda of theories of perception are typically formulated in environment-specific terms, environment-specific contents will best serve the explanatory goals of such theories.[9] The point can be generalized. Given that the pretheoretic explananda of computational theories are typically framed in ordinary language, in terms of publicly accessible objects and properties, and that the content of public language is generally thought to involve essential reference to the subject's physical and social environment, the ascription of wide content to computational states and structures will be appropriate.[10]

A close look at Marr's theory confirms the point. He ascribes wide, 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. In both cases—correlation of a posited structure with a simple distal property in the

[8] See Segal, 1989, 1991, for the most promising account of narrow content in computational vision theory.

[9] Of course, theories do not really have explanatory goals—*theorists* do—but this does not affect the point.

[10] Obvious exceptions are computational theories that attempt to explain our arithmetical abilities.

subject's normal environment or correlation with a disjunctive distal property in the subject's normal environment—the contents ascribed to the structures are wide or environment-specific. The contents so ascribed are determined by the correlations that obtain in the subject's normal environment; correlations that obtain in counterfactual environments are irrelevant.

Some of the structures that Marr posits—for example, individual zero-crossings—do not correlate with any easily characterizable distal property, simple or disjunctive, in the subject's normal environment. Some of their tokenings correlate with distal properties, others appear to be mere artefacts of the imaging process. Marr cautions that such structures are not 'physically meaningful'. For example, he describes zero crossings 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 contents only when an environment-specific distal content (i.e. a wide content) is unavailable.

Causal (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.[11] Some may be tempted to find in Marr's theory support for a causal theory of content. This would be a mistake. I have claimed that in ascribing content Marr looked for salient distal correlates of a structure's tokening in the subject's normal environment. I have avoided saying that a structure represents its normal distal *cause*; Marr certainly made no such claim. Perhaps there is no harm in speaking this way, as long as metaphysicians of content (i.e. philosophers interested in 'the representation relation') do not read too much into such talk. It should be clear that Marr's theory is not committed to a causal theory of content if we consider the case where no *salient* distal correlate (simple or disjunctive) of a structure's tokening can be found. In such cases, 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 function—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. 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.

It should be noted that the structures to which proximal contents are ascribed in Marr's theory may correlate with a salient distal property in a counterfactual environment. For example, in some environment, individual

[11] See, for example, Stampe, 1977; Dretske, 1981; and Fodor, 1990.

zero-crossings may correlate with, or track, physical edges. An interpretation appropriate to this counterfactual environment would ascribe environment-specific (i.e. wide) contents to these structures; it would take them to represent physical edges. The structures correlate with discontinuities in the image in all environments—in the environment we are considering zero-crossings correlate not only with physical edges but also with discontinuities in the image—but a proximal interpretation in this case would serve no explanatory purpose. The ascription of environment-specific distal contents to these structures would enable a Marrian visual theory to explain how the mechanism can recover potentially useful information about its environment. The general point is this: there is nothing necessary about the type of content—narrow or wide—ascribed to a structure in the interpretation of Marr's theory appropriate to the actual world.

Several general conclusions can be drawn from the study of content ascription in Marr's theory. Computational theories are committed to no particular account of content determination, and provide no support for any of the 'naturalistic' theories of content currently popular. *Naturalistic* theories of content attempt to specify, in non-intentional and non-semantic terms, a sufficient condition for a mental representation's having a particular meaning.[12] Such theories should be understood as attempts to explicate the nature of the mental representation relation, hence as metaphysical theses, not as accounts of how content is actually determined in cognitive science. Computational theory provides no support for the idea that there *is* a single representation relation. Most importantly: explanatory considerations govern content ascription in computational models. Different sorts of contents serve different explanatory purposes. Contents are always assigned with an eye to the explanatory goals of the theory.

This account of the role of content in computational psychology allows us to bring together two main theses of Stephen Stich's 1983 book, *From Folk Psychology to Cognitive Science.* Content ascription is context-sensitive, as Stich argued. And computational psychology is essentially formal: computational processes are formally specified; the taxomonic principles of computational theories do not advert to representational content. Stich concludes that content will play no role in mature cognitive science. I part company with Stich in claiming that content can serve the explanatory purposes of computational psychology precisely *because* it is sensitive to important features of the subject's context.

One implication of the foregoing account is that the generalizations of computational psychology will subsume me and my twin-earth doppelganger. Since we are physical duplicates we are computational duplicates. But because explanatory interests are very often quite specific—we might want to know, for example, how an agent's behaviour is related to the local

[12] See Dretske, 1981, 1986; Fodor, 1990; Millikan, 1984; Papineau, 1987; and Stampe, 1977, for various examples of naturalistic theories of content.

potable stuff—intentional interpretations appropriate to me and my twin could be expected to assign different wide contents to our type-identical computational states. Computational theories allow two intuitions, generally thought to be incompatible, to be jointly satisfied: that doppelgangers are identical in *psychologically* relevant respects, and hence should be subsumed under the same psychological generalizations, and that a subject's environment is a determinant of (many of) her mental *contents*.

To recapitulate: computational psychology, I have argued, postulates states and processes that are narrowly individuated, yet makes extensive use of wide or externalist content. This combination is possible only because the states and processes characterized by computational theories do not have their contents essentially. For many this will seem to be an unacceptable consequence of the position I am advocating. But for those who insist that computational states do have their contents essentially there are two options: (1) computational states and processes are narrowly individuated, but then so is content; or (2) both computation and content are externalist. Each of these positions bears a heavy theoretical burden: characterizing an explanatorily useful general notion of narrow content on the one hand, and specifying precisely a notion of externalist computation that fits actual computational practice on the other.[13] Until the burden is discharged, there is no justification for insisting that computationally characterized states have their contents essentially.

The fact, then, that psychology is in the business of explaining not only the behaviour of rational agents but also their intentional states does not imply that the states posited in psychological explanations must be essentially intentional. In fact, it is rather odd to demand that intentional states be explained by appeal to states that are themselves intentional. One would have thought that as thoroughgoing naturalists about the mind, we should hope that intentionality will not turn out to be a fundamental, hence inexplicable, feature of the universe, but rather something that will eventually be explained in terms of more basic, better understood processes. Computational theories have made some progress toward this goal. (Critics such as Searle have interpreted this explanatory progress as the loss of 'intrinsic intentionality'.) The fact that, in making this progress, computational theories move away from the commonsense individuation of mental states, as essentially contentful states, should not be surprising. A principled departure from commonsense schemes is very often a sign of theoretical progress.

Similarly, even if intentional states are individuated externally, in terms

[13] Sections 4 and 5 of Peacocke, 1994, promise 'a positive general account of what is distinctive of content-involving computational explanation' (p. 312). Section 4 is concerned primarily with a defence of the claim that externally individuated states require explanation by externally individuated states, and section 5 with the irreducibility of content-involving explanation to neurophysiology. There is no account here of what is distinctive about *computational* explanation, hence no general account of externalist computation.

of content that makes essential reference to features of the subject's environment, they do not require explanation by states which are themselves externally individuated. Peacocke obscures this fact when he says, in response to a suggestion of mine (Egan, 1992) that individualistic psychological states, when supplemented by assumptions about the organism's normal environment, will provide explanations of organism/environment interaction:

> Perhaps supplementation can help to explain that interaction, but what was wanted was different—it was an explanation of the organism's being in *states whose individuation involves relations to the environment*. (Peacocke, 1994, p. 324, my emphasis)

The states in question are intentional and externalist as *pretheoretically* individuated; but scientific psychology, in developing its explanatory theories of mental phenomena, is not constrained to preserve our pretheoretic way of individuating mental states and processes. The demand by philosophers that it do so exemplifies what Chomsky (1995, p. 28) has called *methodological dualism*, 'the doctrine that in the quest for theoretical understanding, language and mind are to be studied in some manner other than the ways we investigate natural objects'. Needless to say, an analogous demand that the explanatory principles of physics or chemistry respect and preserve the categories of 'folk science' would not be taken seriously.

The arguments of Peacocke and others assume similar unjustified constraints—that intentional facts require intentional explanations, that externalist facts require externalist explanations. Computational theories *appear* to respect these constraints—they appear to be both intentional and externalist—in part because talk about the construction of representations of the environment plays an important role in the informal explication of these theories. Computationally characterized mechanisms do, of course, construct representations of features of the subject's environment. I am not denying this obvious fact, but I have tried to explicate what talk of constructing representations of the environment amounts to *in computational theory itself*. (Some of) the states posited in the theory must be interpretable as representing features of the subject's environment. In certain counterfactual circumstances the same states would not represent those features. They might not represent at all. An intentional state, as characterized by a computational theory, just is a state that is assigned a content in the semantic interpretation appropriate to the actual environment. The states and processes posited in computational models of our representational capacities are neither essentially intentional nor externalist. They do not need to be to explain intentional and externalist facts.[14]

[14] A consequence of the foregoing account is that (*pace* Burge, 1986) individualists need not attempt to construct narrow or 'autonomous' descriptions of behaviour to serve as explananda of narrow psychological theories.

3. An Argument for Narrow Individuation

Psychology attempts to understand the internal states underlying an organism's behaviour and cognitive capacities. Computational psychology, in particular, attempts to characterize the internal mechanisms underlying our cognitive capacities. Of course, to say that psychology characterizes internal states and mechanisms is not to settle the individualism issue, which is not concerned with the *location* of psychological states—externalists do not typically deny that they are in the head, rather than in the environment, or somehow smeared over the conjunction of organism and environment[15]—but with their *individuation*. It is perfectly consistent for internal states and mechanisms to be individuated by reference to an environment (or range of environments). According to an externalist individuation scheme, states and mechanisms that are identical from an internal, physical point of view could count as different states and mechanisms if they are embedded in different environments or are embedded differently in the same environment.

Following Burge (1979), the position that denies that psychological states are individuated by reference to the environment has been called 'individualism', but this name is somewhat misleading. From the point of view of much of theoretical psychology, computational psychology in particular, the boundary between the individual, that is, the organism, and the environment is of no particular interest. Computational processes are usually construed as *modular* processes and characterized independently of the larger system(s) in which they are embedded. Hence, even the *internal* environment is irrelevant to the individuation of a computational mechanism. An adaption of an example from Davies (1991) illustrates the point.[16] 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. Imagine that the auditory system of some actual or imaginary creature contains 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. Now suppose that a particular visex and audex are 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 mathematical functions. The switch will make no difference to the behaviour of the subjects, nor to anything that is going on inside their heads. The visex and audex are computationally identical, despite the difference in their normal internal environments. Visex and

[15] Although Chalmers and Clark, 1998, argue that minds are realized, in part, by aspects of the environment.

[16] Davies, 1991, uses the example in support of an *anti-individualist* construal of computational mechanisms.

audex, from the computational point of view, are the same mechanism. The surrounding organism is just so much environment.[17]

It is true, of course, that in its normal (i.e. original) internal environment the visex computes a representation of depth from disparity. More precisely, the computational vision theory that describes the visex interprets it as computing a function defined on the visual domain. But the content assigned to states of the device by an interpretation that is appropriate to its normal environment is not an essential property of the device as computationally characterized. In a different internal environment—embedded in the auditory system of some other creature—it computes a different intentionally characterized function, and hence will be assigned different content.

It follows from this that the states characterized by a computational theory of vision are not essentially *visual* states. Some commentators have taken this consequence to be a fatal objection to my account of computational vision theory (see Butler, 1996, and Shapiro, 1997, who have taken it to be a virtual reductio of my position). But I don't see why it should be regarded as an objection at all. Visual states are a species of intentional state—they have representational content. A visual state is a (certain sort of) representation of (certain) features of the environment. If computationally characterized states are not essentially intentional, then they will not be essentially visual either. But there is no reason to insist that our visual abilities must be explained by reference to internal states that are themselves *essentially* visual. A visual state, according to computational vision theory, just is a state that is assigned a visual content in the interpretation appropriate to the actual environment, nothing more.

It is, of course, consistent with individualism to maintain that the visex and audex are different computational mechanisms because they are normally embedded in different internal environments. The individualist need only hold that the *external* environment is irrelevant to psychological individuation. In denying that the organism/environment distinction has any individuative significance within computational theory,[18] I am arguing for a *narrower* individuation scheme that is required by individualism. The narrow/wide distinction defines a partial ordering on which individuative principles in psychology can be compared, rather than a simple dichotomy.

[17] Compare what Chomsky says about an *I-language*, a similarly narrowly individuated object:

> It is only by virtue of its integration into such performance systems [the performance systems involved in language comprehension and speech production] that this brain state qualifies as a language. Some other organism might, in principle, have the same I-language (brain state) as Peter, but embedded in performance systems that use it for locomotion. (Chomsky 1992, p. 213)

[18] I am not claiming that the internal environment is irrelevant for determining an appropriate interpretation.

Very roughly, the wider the individuative scheme, the greater the range of relational or contextual properties of an internal state (mechanism, process) that are relevant for its type-identification. A scheme that takes features of a mechanism's embedding within the organism to be relevant to its type-identity is wider than one that denies that these features help determine the type of mechanism that it is, but narrower than a scheme that, in addition, takes features of the organism's historical, social, and environmental context to play an individuative role.

I want to suggest that, where the primary goal is to understand the mechanisms and processes underlying the behaviour of a complex system, there is a presumption in favour of (relatively) narrow individuation in science: very simply, the narrower the individuative scheme, the greater the *scope* of the theory's generalizations. *Generality* is a desideratum of any explanatory theory (not, of course, the only desideratum), and so, other things being equal, (relatively) narrow individuative principles are preferable to principles that build in additional aspects of the environment or context.

The controversy over twin-earth examples illustrates why this is so. The prime motivation behind the attempt to specify some notion of narrow content shared by doppelgangers is the intuition that they are identical in (at least some) psychological respects. Their behaviour, their dispositions to behave, their cognitive capacities and abilities are the same. These commonalities are obscured by our commonsense individuative schemes—inasmuch as these schemes are externalist the twins' behaviour is *not* the same. Nonetheless, the commonalities are there to be uncovered, explicitly characterized, and explained. A psychology that preserves the commonsense scheme and individuates the twins' mental states in terms of their wide content, which is sensitive to differences in the environments which have no effect on the twins' physical states, risks forgoing a deeper understanding of the springs of behaviour.

Twin-earth cases are a philosophers' fiction. But if we were actually to discover, on a far-away planet, creatures who behaved (to all appearances) exactly as we do, we would have no trouble understanding the impulse, among the more scientifically inclined, to explain the commonalities between us and them by positing underlying states that abstract away from the contextual differences (history, environment, etc.) between our mental states and theirs. Of course, this explanatory project might not pan out. Our underlying psychologies could be very different despite the (apparent) similarities in our behaviour. Or, for any number of reasons, any deeper affinities might be very difficult to characterize. (Theorists may attempt to pick out shared underlying states using descriptive devices such as 'the state that is shared by all individuals of whom one of some specified list of relational descriptions is true'. The resulting 'theory' is likely to be of dubious explanatory value. What is wanted is a characterization of the underlying state that is independent of the relational properties that the explanatory principles of the theory prescind from.) The important point is that the strategy motivating the construction of narrow taxonomies is impeccable, whatever the ultimate fate of the theories it produces.[19]

The same strategy motivates the construction of computational models of mind. It is an interesting (and explanatory) fact about a system that it computes the same function as a class of well-understood mathematical devices. There are generalizations to be captured at the computational level of description that will be missed if features of the system's (internal or external) environmental context are built in to the individuative principles of the theory. These relational facts are relevant for the ascription of *content* to the posited states and processes, and so the explanations of intentional facts yielded by the theory will take them into account, as explained above. They help determine the appropriate semantic interpretation, which is necessary to explain the device's representational, as opposed to purely computational, abilities. (What the device represents, though not what mathematical function it computes, does depend upon its environmental context.) Nothing is to be gained, and much explanatory potential would be lost, by collapsing computational individuation and content individuation. Thus, a computational theory would treat the visex and the audex as fundamentally the same device, prescinding from their normal (internal and external) environments.

We need not rely on a thought experiment to illustrate the point. Marr describes a component of early visual processing responsible for the initial filtering of the image. Externalists like to point out that Marr's theory characterizes components of the visual system in terms of what they do, their task or purpose (see, for example, Shapiro, 1993, 1997). This is true, but the interesting question is how the theory specifies the relevant task. The task of a computational mechanism is to compute a certain mathematical function; the initial visual filter, for example, computes the Laplacean convolved with a Gaussian. The device computes this mathematical function whether it is part of an auditory or a visual system, in other words, independently of the environment in which it is embedded. In fact, it is likely that each sensory modality has one of the same computational devices—since the device just computes a curve-smoothing function. It's a real-life visex/audex!

A theory whose individuative principles prescind as far as possible from features of a system's context does not thereby ignore or underestimate the system's environment as a determinant of its behaviour. Rather, if a theory with a (relatively) narrow individuative scheme is to achieve predictive and explanatory adequacy, the theorist is forced to separate and independently specify aspects of the context that contribute to the system's behaviour. Understanding complex behaviour—whether it be the movement of a body on an inclined plane, the fall of a leaf during a windstorm, or the intelligent

[19] Readers may notice an affinity between the point advanced here and McGinn's (1991) distinction between *powers* (which are taxonomic) and *parameters* (which are not). McGinn, however, takes the distinction to apply to aspects of *content*. On the account advanced here, content is not a power.

behavior of a rational agent—as the result of the interaction of a number of independently specifiable variables is a hallmark of post-Galilean science. (This explanatory strategy manifests itself in computational cognitive science in the *principle of modular design* (Marr, 1982, p. 102), which enjoins the theorist to, wherever possible, characterize a complex process as the outcome of independently specifiable operations.) Sometimes the number of variables is too great, or the interactions too complex, to allow the theorist to actually predict the system's behaviour (e.g. the precise path of a leaf's fall), but in such a case we have an 'engineering problem', not a failure of explanatory strategy. When it succeeds, the strategy allows us to achieve a generality of understanding. We can see not only why the system behaves as it does, but also how it would have behaved differently had any number of its relational properties been different. This is a theoretical virtue that externalists seem to have overlooked.

Department of Philosophy
Rutgers University

References

Block, N. 1986: Advertisement for a Semantics for Psychology. *Midwest Studies in Philosophy*, 10, 614–78.

Burge, T. 1979: Individualism and the Mental. *Midwest Studies in Philosophy*, 4, 73–121.

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

Butler, K. 1996: Content, Computation, and Individualism in Vision Theory. *Analysis*, 56, 146–54.

Chalmers, D. and Clark, A. 1998: The Extended Mind. *Analyst*, Preprint 32.

Chomsky, N. 1992: Explaining Language Use. *Philosophical Topics*, 20, 205–31.

Chomsky, N. 1995: Language and Nature. *Mind*, 104, 1–61.

Cummins, R. 1989: *Meaning and Mental Representation*. Cambridge, MA: MIT Press.

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

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

Dretske, F. 1986: Misrepresentation. In R. Bogdan (ed.), *Belief: Form, Content, and Function*. Oxford University Press.

Dretske, F. 1995: *Naturalizing the Mind*. Cambridge, MA: MIT Press.

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

Egan, F. 1992: Individualism, Computation, and Perceptual Content. *Mind*, 101, 443–59.

Egan, F. 1995: Computation and Content. *Philosophical Review*, 104, 181–203.

Fodor, J.A. 1975: *The Language of Thought*. New York: Crowell.

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

Fodor, J.A. 1987: *Psychosemantics*. Cambridge, MA: MIT Press.

Fodor, J.A. 1990: *A Theory of Content and Other Essays*. Cambridge, MA: MIT Press.

Marr, D. 1982: *Vision*. New York: Freeman.

McGinn, C. 1991: Conceptual Causation: Some Elementary Reflections. *Mind*, 100, 573–86.

Millikan, R. 1984: *Language, Thought, and Other Biological Categories*. Cambridge, MA: MIT Press.

Papineau, D. 1987: *Reality and Representation*. Oxford University Press.

Papineau, D. 1993: *Philosophical Naturalism*. Oxford: Blackwell.

Peacocke, C. 1994: Content, Computation, and Externalism. *Mind and Language* 9, 303–35.

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

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

Segal, G. 1991: Defence of a Reasonable Individualism. *Mind*, 100, 485–93.

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

Shapiro, L. 1997: A Clearer Vision. *Philosophy of Science*, 64, 131–53.

Stampe, D. 1977: Toward a Causal Theory of Linguistic Representation. *Midwest Studies in Philosophy*, vol. 2.

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

Tye, M. 1995: *Ten Problems of Consciousness*. Cambridge, MA: MIT Press.

Wilson, R. 1994: Wide Computationalism. *Mind*, 103, 351–72.

Wilson, R. 1995: *Cartesian Psychology and Physical Minds*. Cambridge University Press.