1

Is There a Role for Representational Content in Scientific Psychology?

FRANCES EGAN

Steve Stich used to be an eliminativist. As far as I can tell, he renounced eliminativism about the time that he moved from the west to the east pole. Stich was right to reject eliminativism, though I am not convinced that he rejected it for the right reasons. Stich 1983 contains a comprehensive attack on representational content, a central feature of both folk psychology and the Representational Theory of Mind, the leading philosophical construal of scientific psychology. Stich’s current position on the role of content in psychological explanation is not entirely clear. One of my aims in this chapter is simply to invite Stich to clarify his views on representational content; the question that forms the title of this chapter is therefore addressed directly to Stich. I begin by sketching his original anti-content argument. I then trace some later developments in his thinking about content. I argue that content does play an important role in scientific psychology, for precisely the reasons that Stich identified in his original argument against content. I conclude with some general remarks on eliminativism.

I

Stich characterizes four distinct claims that eliminativists are inclined to make:

1 ‘Belief’, ‘desire’, and other familiar intentional-state expressions are among the theoretical terms of a commonsense theory of the mind. This theory is often called folk psychology.

2 Folk psychology is a seriously mistaken theory. Many of the claims it makes about the states and processes that give rise to behavior, and many of the presuppositions of these claims, are false.

3 A mature science that explains how the mind/brain works and how it produces the behavior we observe will not refer to the commonsense intentional states and processes invoked by folk psychology. Beliefs, desires, and the rest will not be part of the ontology of a mature scientific psychology.

4 The intentional states of commonsense psychology do not exist. (Stich and Ravenscroft 1994: 116)
Claim (1) is simply a presupposition of eliminativism. Realists about beliefs and desires typically endorse it too. We will call claim (3) weak eliminativism (following Stich 1992: 245), and a commitment to either (2) or (4) strong eliminativism. (2) and (4) are, of course, distinct claims, but we will note their differences only where relevant. Stich 1983 endorsed all three eliminativist claims. He also assumed there that (3) and (4) follow from (2).

Stich’s 1983 argument for eliminativism was two-pronged. In the latter part of the book, he argued that folk psychology makes substantial assumptions about the cognitive architecture underlying our cognitive capacities and behavioral repertoire. He claimed, for example, that folk psychology is committed to the idea that a single memory system suberves verbal and non-verbal behavior (p. 231), and that belief organization and storage is modular. There is good reason, the argument continues, to suspect that these assumptions are false. I argue in Egan 1995a that folk psychology involves no substantive commitments about architecture or cognitive processing, and so the falsity of these architectural assumptions would not threaten it. I shall not defend this claim here, though the issue will come up again in the last section.

My concern here is with the other prong of Stich’s 1983 argument for eliminativism, that is, his case against representational content. First, some stage-setting. Folk psychological predictions and explanations of behavior appeal to content-specific beliefs and desires. For example, it is my belief that there is beer in the refrigerator together with the content-appropriate desire (to drink a beer, or perhaps just to drink something cold), that explains my going into the kitchen and getting a beer. Appealing to my belief that there is beer at the local bar or my desire to win the lottery fails to provide an explanation of my beer-fetching behavior. Moreover, this behavior is rational just to the extent that it is caused by content-appropriate beliefs and desires.

The case against content in Stich 1983 is a tour de force. Stich argues persuasively that content ascriptions are both vague and context-sensitive. For any given predicate of the form ‘believes that p’ there will be many contexts where there is simply no saying whether it applies or not, and hence it will often be unclear whether a generalization that invokes such a predicate applies to a given subject. Appeals to content are also observer-relative. As Stich puts it, “To believe that p is to be in a belief state similar to the one underlying our own sincere assertion of ‘p.’” (136). Moreover, appeals to content often presuppose both ideological similarity and reference similarity. Two beliefs are ideologically similar if and only if they are embedded in similar doxastic networks. Suppose, for example, that two subjects both say “Senator Smith is a liberal.” Whether or not we would be inclined to attribute the same belief – that is, a belief with the same content – to the two subjects depends on whether their other beliefs involving the concept liberal are similar. Two beliefs are reference similar if and only if the terms subjects use to express the beliefs have the same referent. Subjects using the vocable ‘granite’ to refer to different substances would express different beliefs when they say “granite counter-tops are durable.”

Let us call these various properties of content – vagueness, context-sensitivity, observer-relativity, the tendency to presuppose ideological and reference similarity – the R (for ‘relativity’) properties. The problem, Stich argues, is that a taxonomy that has the R properties will impose a more fine-grained individuative scheme than is appropriate for use in scientific psychology. In particular, a taxonomy that has the R properties will not respect the Autonomy Principle, which holds that “any state or property properly invoked in a...
psychological explanation should supervene on the current, internal, physical state of the organism.” Stich argues (1983: 167–9) that systematic explanations of an organism’s behavior of the sort that psychologists seek to provide – as opposed to those sought by, say, social historians or biographers – will apply equally to an organism’s physical duplicate. Such explanations should invoke only narrow states and properties shared by all duplicates; in particular, they should invoke only narrow causal role. But states individuated in part by their content, as beliefs and desires are, build in various features of the subject’s historical, environmental, and social context, features that are irrelevant for scientific psychology.

Stich offers what he calls the ‘replacement argument’ in support of the autonomy principle (1983: 165). Since physical duplicates would behave identically in all contexts, and since psychology is the science that aspires to explain behavior, any states or properties not shared by duplicates must be irrelevant to psychology. Of course, there are some behavioral descriptions that do not apply equally to duplicates. I can sell you my car; a physical duplicate of me cannot. Stich proposes that psychology restrict itself to ‘autonomous behavior descriptions’, that is, descriptions that satisfy the following condition: if they apply to an organism in a given setting then they would apply to any replica of the organism in that setting. Scientific psychology should aim to purge itself of language that builds in historical or contextual features. It should purge itself, therefore, of representational content.

So, to summarize: one important strand in Stich’s 1983 case for eliminativism, which we will call the 1983 anti-content argument, is the following:

1 Content has the $R$ properties. (It is vague, context-sensitive, presupposes various dimensions of similarity, etc.)
2 States and properties invoked in a scientific psychology should supervene on the current physical state of the organism (the Autonomy Principle).
3 Properties with the $R$ properties do not supervene on the current physical state of the organism, and so a taxonomy based on such properties will violate the Autonomy Principle.
4 Therefore, content should not be invoked in a scientific psychology.

II

Stich’s original argument concerned ordinary content ascribed in folk psychological predictions and explanations of behavior. In Stich 1991, he argued that narrow content – content that prescinds from the subject’s historical, environmental, and social context, and hence is shared by physical duplicates – is nonetheless still too vague and context-sensitive (in other words, it still has many of the $R$ properties) to serve in a scientific psychology. In particular, he argued, narrow content is ill-suited to play a role in computational models of mind. Such models individuate mental states in terms of their narrow causal role but, as numerous examples show, narrow content does not track narrow causal role.

Summing up, Stich says “the categories of a narrow content taxonomy are simply the categories of a broad content taxonomy extended to meet the demands of the principle of autonomy. But the broad content taxonomy of commonsense psychology is too vague, too context-sensitive and too unstable to use in a serious scientific theory. Narrow content
inherits all of these deficits” (1991: 250). Clearly, at this point, Stich still believed that content, broad or narrow, is not suitable for science. And he still endorsed weak eliminativism – scientific psychology will not invoke commonsense intentional states because these states are individuated by content. The argument, which does not rely on the Autonomy Principle, looks something like this:

1 Content has the \( R^* \) properties. (Where \( R^* \) properties are the proper subset of the \( R \) properties that meet the demands of the Autonomy Principle, including, vagueness, context-sensitivity, and instability.)
2 Properties with the \( R^* \) properties (vagueness, context-sensitivity, etc.) are not suitable for science.
3 Therefore, content should not be invoked in a scientific psychology.

We will call this the 1991 anti-content argument.

By 1996 Stich’s view has changed considerably. In 1996c, he says “being invoked in a successful science is all that it takes to render a property scientifically legitimate. On my view, the jury is still out on the question of whether successful science can be constructed using intentional categories” (1996c: 199). But one would have thought that the jury had already handed down its verdict. Has content somehow been rehabilitated? Let’s consider some recent developments in Stich’s thinking that might account for his change of mind:

(1) The so-called “naturalization project” is the attempt to specify, in a non-intentional and non-semantic vocabulary, sufficient conditions for a mental state’s meaning what it does. Most, if not all, attempts at naturalization have failed to meet these stringent requirements while characterizing something that looks sufficiently like representational content. Stich argues persuasively (in Stich 1992 and Stich and Lawrence 1994) that the failure of the naturalization project would not impugn intentional content. Content, whatever its other failings, does not need to be naturalized. But this conclusion should give Stich no reason to reconsider whether content is fit for use in scientific psychology. His case against content – in particular, the 1983 and 1991 anti-content arguments – does not depend upon metaphysical or philosophical considerations of the sort that vex those engaged in the naturalization project.

(2) Stich has become convinced that even if folk psychology is a seriously mistaken theory (claim 2 on p. 14), it does not follow either that scientific psychology will find no use for intentional categories (claim 3), or that beliefs and desires do not exist (claim 4). To think otherwise, as Stich once did, he argues in 1996b, is to make some dubious assumptions about the reference of theoretical terms. But the failure of the inference from claim (1) to (3) and (4) does not undermine the 1983 or the 1991 anti-content arguments, neither of which depends upon any assumptions about reference. The failure of the inference provides no reason to doubt that content has the \( R \)-properties, and so no reason to be optimistic that successful psychology will invoke intentional states.

(3) The 1991 anti-content argument does not depend upon the Autonomy Principle. Still, it is worth noting that Stich is no longer willing to endorse the principle. In both 1991 and 1996b he holds the Autonomy Principle at arm’s length, claiming:
There is, to put it mildly, considerable controversy surrounding this thesis. Some writers, myself included (but I was younger and much more naive at the time), have claimed that it is intuitively obvious . . . Others have tried to defend the thesis by deducing it from other, perhaps less controversial, metaphysical doctrines; still others have claimed that it is simply false. (1996b: 23)

He goes on to say: “But if it is not clear whether the thesis is defensible, it is clear that if the thesis is accepted then . . . folk psychology is in trouble” (1996b: 23). The Autonomy Principle requires that psychological properties supervene on the current physical state of the subject, and so it clearly prohibits individuating psychological states in terms of their broad content, as folk psychology appears to do. Rejecting the Autonomy Principle would let broad content back into the picture, if there were not independent reason to think that content – broad or narrow – is also vague, context-sensitive, and unstable. There is no obvious reason for Stich to repudiate his 1991 anti-content argument, which relies on the claim that content has the \( R^* \) properties and hence is not suitable for use in a scientific psychology.8

In short, recent developments in Stich’s thinking do not account for his apparent change of mind about content. I will wrap up this section with a question for Stich: are there (still) principled reasons why scientific psychology should avoid a commitment to representational content?

III

Here is what I will argue:

1 Content has the \( R \) properties.
2 A qualified version of the Autonomy Principle is true. In particular, a theory concerned primarily with characterizing the mechanisms underlying our cognitive capacities should employ taxonomic principles that prescind from the mechanism’s normal context.
3 Therefore, a theory concerned primarily to characterize the mechanisms underlying our cognitive capacities should not individuate by properties with the \( R \) properties (e.g., content) – it will violate the Qualified Autonomy Principle.
4 The theories that are in the business of characterizing the mechanisms underlying our cognitive capacities – computational cognitive theories – in fact, do not individuate by content.
5 Content nonetheless plays an important explanatory role in such theories, and it is able to play this role in part because it has the \( R \) properties.

I will not defend (1); I take it to have been established by Stich 1983. Content is vague, observer-relative, context-sensitive; it presupposes reference and ideological similarity. My argument for claim (4) will be quite brief, as I have defended it at length elsewhere. Turning, then, to the argument for (2). Recall that the Autonomy Principle holds that the only properties to be invoked in a scientific psychology are narrow properties that supervene on the current physical state of the organism. When the principle is appropri-
ately qualified it has a compelling rationale. If a theory is concerned primarily to characterize the mechanisms and processes underlying the behavior and capacities of a complex system, then relatively narrow taxonomies are better for the following reason: the narrower the individuative scheme— that is, the greater the range of contextual properties it prescinds from—the greater the scope of the theory’s generalizations. Narrow taxonomies maximize generality. Hence, the Qualified Autonomy Principle (henceforth, QAP) holds that a psychological theory concerned primarily to characterize the mechanisms underlying our cognitive capacities should employ taxonomic principles that prescind from features of the mechanism’s normal context.

Consider the inhabitants of Twin Earth. According to Putnam’s (1975) myth, their cognitive capacities and dispositions to behavior are the same as ours. Characterizing the underlying commonalities between ourselves and Twin Earthlings, rather than obscuring such commonalities by building contextual features into our taxonomies, provides a basis for explaining and predicting our behavior, and theirs, in a wide range of counterfactual circumstances. This strategy does not ignore or downplay the subject’s environment as a determinant of her behavior. It simply requires that environmental and other contextual determinants of behavior be specified as independent variables.

QAP is in one sense stronger than Stich’s original principle. The states and properties invoked by theories that purport to characterize the mechanisms responsible for cognitive capacities will typically supervene not just on the current physical state of the whole subject, but more narrowly on the current physical state of the mechanism itself. The theorist is directed to characterize the relevant mechanisms and processes independently of the larger systems in which they are embedded, prescinding from details of both external and internal (i.e., intra-organism) environment. The resulting theories will be radically internalist; the boundary of the subject’s skin will have no particular individuative significance.

QAP is a special case of a more general principle about how to understand and explain complex systems. As noted above, taxonomies that prescind from contextual features, requiring that they be specified as independent parameters, provide a basis for predicting and explaining the system’s behavior in a wide range of counterfactual circumstances. The presumption in favor of narrow taxonomies applies only to theories in the business of explaining how complex systems work. It does not apply to “historical” theories, such as evolutionary biology and geology, whose explanatory goals are somewhat different.

Thus, a theory that aimed to explain how circulation is possible might subsume human and Martian hearts under the same kind if they worked the same way, whereas evolutionary biology, with an interest in the specific origins of biological mechanisms on earth, type-identifies only homologous organs.

In another sense, then, QAP is weaker than Stich’s Autonomy Principle. It does not require scientific psychology to restrict itself to narrow taxonomies. Not all of psychology is concerned to characterize the mechanisms underlying cognitive capacities and behavior. Computational psychology clearly is, but other branches of scientific psychology have other goals. Developmental psychology, for example, is concerned to characterize the particular stages in a human child’s cognitive development, and so we would not expect it to respect QAP.

It follows from (1) and (2) that a theory concerned primarily to characterize the mechanisms underlying our cognitive capacities should not individuate psychological
states and processes by reference to content (i.e., 3 in my argument). Physical duplicates in different environmental and social contexts may not share intentional content. A computational cognitive theory that individuated the mechanisms it characterizes by reference to content is therefore likely to violate QAP.

It might be argued that only ordinary, broad content violates QAP. It is open to a psychological theory to specify its own special type of narrow content and individuate psychological states and processes in terms of it. The specified narrow content might even be free of the other R properties that Stich 1983 and 1991 claimed make content unsuitable for use in a science. It might not be vague, context-sensitive and unstable. But we won’t know until we see it.12

I claim that computational cognitive theories do not individuate in terms of any type of content (4 in my argument). Nonetheless, content does play an important role in computational models. It is able to play this explanatory role precisely because it is both interest relative and context sensitive (5 in my argument). I defend these claims in the next section.

IV

I have argued in a series of papers (1992, 1995b, 1999) that computational cognitive theories do not individuate the states and processes they characterize by reference to intentional content. A given computational state might, in some counterfactual circumstances, have a different semantic content, or no content at all, and nonetheless be the same computational state.

Disputes about whether or not computational theories individuate the states they characterize in semantic terms – by their content – turn on how the level of description that David Marr called the theory of the computation should be interpreted. The theory of the computation provides a canonical description of the function(s) computed by a computational mechanism. It specifies what the device does.13 By a “canonical description,” I mean the characterization that is decisive for questions of individuation or taxonomy. In Egan 1995b, I argued that the canonical description of the function computed by a computationally characterized mechanism is a mathematical description. An example from Marr’s own theory of vision illustrates the point. Marr (1982) describes a component of early visual processing responsible for the initial filtering of the retinal image. Although there are many ways to informally describe what this filter does, Marr is careful to point out that the theoretically important characterization, from a computational point of view, is a mathematical characterization: the device computes the Laplacean convolved with the Gaussian (Marr 1982: 337). As it happens, it takes as input light intensity values at points in the retinal image and calculates the rate of change of intensity over the image. But as far as the computational characterization of the device is concerned, it does not matter that input values represent light intensities and output values the rate of change of light intensity. The computational theory characterizes the visual filter as a member of a well-understood class of mathematical devices that have nothing to do with the transduction of light.14
The computational characterization abstracts away from the visual function the device performs in its normal (external and internal) environment, and hence from visual contents normally attributed to the internal states and structures characterized by the theory. The very same mechanism might be found in the auditory system computing the same curve-smoothing function but defined over acoustical properties. Following Martin Davies (1991), we can call the mechanism a visex when it is embedded in the visual system, and an audex when it is embedded in the auditory system. Davies thought he was describing a far-fetched thought experiment, and he thought it obvious that psychology would treat visex and audex as type-distinct computational mechanisms. But the example is not far-fetched at all. For all we know, there might be a mechanism that computes the same curve-smoothing function in one of our other sensory systems. Provided that this mechanism uses the same algorithm to compute this function, a computational theorist would treat it, and the device described by Marr, as type identical. The mathematical characterization that is crucial for type-identifying the device as a computational mechanism prescinds from the device’s normal environment. In other words, it respects the Qualified Autonomy Principle. In doing so, this characterization provides the basis for predicting and explaining how the mechanism will behave whatever its (internal and external) environment.

I turn, finally, to the defense of claim 5 of my argument. Representational content does not play an individuative role in computational cognitive theories; but it does play an important explanatory role. It is able to play this role in part because it has the R properties, in particular, because it is both interest relative and sensitive to various aspects of the subject’s context, and hence is not a property that respects the Qualified Autonomy Principle.

A semantic interpretation of a computational system is given by an interpretation function that specifies a mapping between equivalence classes of physical states of the system and elements of some represented domain. To interpret a device as a visual system is to specify a mapping between states of the device and tokenings of visible properties such as changes of depth in the scene; to interpret a device as a parser is to specify a mapping between states of the device and syntactic items such as noun phrases or verb phrases. The specified states of the device are thus construed, under the interpretation, as representations of changes in depth, or of noun phrases.

What does the semantic interpretation add to the computational characterization of the device, which construes it as computing a particular mathematical function? I have argued that it does not serve an individuative function. The semantic interpretation is necessary to explain how the abstractly characterized process, in a certain context (say, when situated in a certain external environment, or connected to certain performance systems) constitutes the exercise of a cognitive capacity, such as perceiving the depth of objects in the scene, or parsing a sentence. The questions that define a psychological theory’s domain are typically couched in intentional terms. For example, we want a theory of vision to tell us how the visual system can detect three dimensional structure from information contained in two dimensional images. We want a theory of language understanding to explain how the subject can recover the meaning of an acoustical signal. It is only under an interpretation of some of the states of the computationally characterized
mechanisms as representations of visible distal properties (changes of depth), or as representations of syntactical categories (noun phrase), that the processes characterized in abstract mathematical terms by the theory are revealed as vision, or as parsing. Unless some of the internal states described by the theory can be interpreted as representations of visible distal properties, the computational visual theorist can make no claim to have described a mechanism that enables an organism to see. So, the semantic interpretation forms a bridge between the intentionally characterized explananda of the theory and the abstract, mathematical characterization of the mechanism that constitutes the explanatory core of a computational theory.

We can specify the cognitive (as opposed to the mathematical) function subserved by a computational mechanism only by considering how it is embedded in the surrounding environment, including the internal environment. In some counterfactual environments, the mechanism might fail to compute the specified cognitive function. The point can be most clearly seen for perceptual capacities. In computational models of perception, the content ascribed to internal states of the device will be determined by the distal properties tracked by these internal states. For example, the structures that Marr calls edges are tokened in the presence of a disjunctive distal property, namely, a change in depth, surface orientation, illumination, or reflectance. So they represent a salient physical boundary of a general sort. But these structures cannot be expected to track, hence to represent, this property in every possible environment. In some weird counterfactual environments they might track no salient or easily characterizable property; in such circumstances they would represent no distal property. If a Marrian visual system were somehow to appear in such an environment (say, as a result of a genetic accident or an experiment by IBM) it would not enable its possessor to see. The mechanism would still compute the same mathematical function, but computing this function, in this environment, would not enable it to detect a salient or useful property of that environment.15

The point I wish to emphasize is that constraints on content ascription are both interest relative and context-sensitive. That they are interest relative can be seen by reflecting on the oft-noted fact that interpretation is not unique. The internal states and structures of a given mechanism may co-vary with any number of properties, and hence support multiple interpretations. But the semantic interpretation assigned to the states of a mechanism in a computational cognitive model reflects our interest in explaining the organism’s success at some particular cognitive task, in this case, recovering the 3-D structure of the scene. Opponents of computationalism, such as John Searle (1991), have appealed to the interest relativity of interpretation to argue that computational models are not genuinely explanatory. But the interest relativity noted above does not undermine a computational vision theorist’s claim to have explained how the mechanism performs a visual task, even in the unlikely event that the mechanism’s states also happen to track the New York Yankees’ batting averages. It does not follow that representation in computationally characterized systems is purely interest relative, that such systems do not, as a matter of objective fact, really represent, that ‘anything goes’. Some of the constraints on interpretation are objective. It is an objective fact that the mechanism, in its normal environment, goes into states that track, or co-vary with, changes in depth.

This brings us to the second point: that constraints on content ascription are context sensitive. A given mechanism will subserve a cognitive function, such as vision, only in
The content ascribed to states of the mechanism is therefore likely to be expected to be environment-specific (that is, broad), enabling an explanation of how the mechanism succeeds at recovering information about its normal environment. The semantic interpretation specifies the properties tracked by states of the mechanism when it is functioning properly in its normal (internal and external) environment. Change the environment – populate it with objects that behave very differently, or change the laws of optics – and the interpretation will no longer be supported.

I have been leaning heavily on theories of perception to make the point. But the point has wider application. Suppose that we had a complete computational account of our cognitive capacities. Such an account would specify the mechanisms underlying our perceptual capacities, language understanding and production, reasoning, and so on, and it would provide a basis for the explanation and prediction of behavior. The theory would respect the Qualified Autonomy Principle and hence would apply to our physical duplicates in counterfactual environments. It would subsume our computational mechanisms and theirs under the same abstract mathematical description, and it would specify the basic causal operations of our type-identical mechanisms. But given that the explananda of psychological theories are expressed in ordinary language, in terms of publicly accessible objects and properties, and that the content of public language involves essential reference to the subject’s physical and social environment, we should expect the semantic interpretation that enables the theory to address these explananda to be specific to our world. Consequently, semantic interpretations appropriate to me and my twin-earth counterpart would assign different broad contents to our type-identical computational states.

If this account is right, then Stich’s claim that psychology should stick to autonomous behavioral descriptions is wrong. The mathematical characterization that is decisive for individuating a computational mechanism is autonomous – it will respect the Qualified Autonomy Principle and so apply to any physical duplicate. But the semantic interpretation that accompanies the mathematical characterization is not. It will make reference to the mechanism’s historical and contextual features.

In summary, then, representational content can serve the explanatory purposes of scientific psychology precisely because it is interest-relative and sensitive to the subject’s context. Having the R properties, as Stich argued so persuasively it does in the 1983 book, does not disqualify content from playing an important role in scientific psychology.

V

I shall conclude with some remarks on eliminativism:

1. I have argued that scientific psychology, computational cognitive science in particular, invokes content in its explanations of our cognitive capacities. Does it follow that it invokes beliefs and desires? If it does, then weak eliminativism – the claim that beliefs, desires, and the other propositional attitudes invoked by folk psychology will not be part of the ontology of a mature scientific psychology – is false.

   The conspicuous successes of computational psychology have been in characterizing highly modularized, informationally encapsulated processes such as those responsible for
early vision and syntactic and phonological processing. The states posited by these theories do not have the complex functional roles characteristic of the propositional attitudes, including, typically, accessibility to consciousness. Even if, as I have argued, many of these states are assigned content in computational accounts, there is little reason to identify these states with beliefs and desires. They are paradigmatic examples of so-called subdoxastic states.

The states involved in domain-general processes, such as the processes underlying decision making and rational revision of belief, would make better candidates for identification with propositional attitudes, but these processes have so far resisted computational treatment. Their intractability is due in part to the fact that general constraints on the information that might be relevant to decision-making and belief revision are difficult, if not impossible, to specify. As Fodor 2000 notes, just about anything might be relevant to such processes. He concludes that the prospects for a computational treatment are dim.

In any event, nothing in current computational psychology supports the claim that our best explanations of how the mind/brain works will invoke beliefs and desires. And I think there is a principled reason to be skeptical. Propositional attitudes find their home in personal level psychology, where the goal is to predict and explain how subjects behave in their interactions with each other and the world. Computational psychology, on the other hand, is concerned with explaining the mechanisms underlying subjects' cognitive capacities, which typically requires decomposing these capacities into their functional (i.e., sub-personal) components. Functional decomposition is the dominant explanatory strategy of computational psychology.

Moreover, folk psychology is concerned with more than prediction and explanation. Notions such as belief, desire, and intention are central to our conceptions of personal identity and the self, which underpin a whole host of social practices, including assigning moral and legal responsibility to agents. There is no reason to think that these rich and complex notions will serve the rather austere explanatory purposes of computational psychology.

Of course, it is always possible to expand the list of propositional attitudes to include explanatory constructs from scientific psychology. We could say, following Chomsky (1980), that subjects cognize (rather than know or believe) the contents attributed in computational models, where ‘cognize’ is a technical term introduced to refer to whatever relation subjects in fact bear to these contents. In other words, we could claim that scientific psychology has discovered new kinds of propositional attitudes. But calling cognizing a propositional attitude doesn’t make it one. Beliefs, desires, intentions, fears, and the other standard propositional attitudes form something like a natural kind. They have very complex functional roles; they are inferentially promiscuous; they are typically accessible to consciousness, etc. Cognitions, in Chomsky’s sense, have none of these properties. Admitting them, or similarly attenuated counterparts of beliefs and desires, into the club turns the notion of propositional attitude into a motley with no clear rationale.

But other branches of scientific psychology do invoke beliefs and desires, so weak eliminativism is false. Attribution theory is the branch of social psychology that studies the perceived causes of behavior. (See Heider 1958 for the classic statement of attribution theory, and Weiner 1990 for a more recent survey of the attribution literature.) Often these perceived causes include beliefs, desires, hopes, fears, and so on. For example, an
agent’s failure to expend the amount of effort required to secure a goal may be attributed to his fear that he will fail; a world class athlete’s Herculean efforts in the face of adversity might be attributed, in part, to her belief that she is the best at her sport. Developmental psychology attempts to characterize the commitments that infra-linguistic humans bring to their interactions with the world. Developmental theories attribute to infants beliefs and expectations about how objects move in space (see, for example Spelke 1990).

There are striking differences between the branches of scientific psychology that do invoke propositional attitudes – attribution theory and developmental psychology, to name just two – and computational psychology, which I claim does not. Computational psychology is in the business of explaining how the mechanisms underlying our cognitive capacities work. A computational account typically begins with a problem posed at the personal level – for example, how does the organism recover the 3-D structure of the scene? – but it quickly abandons the personal stance in favor of sub-personal mechanisms that interact to produce complex behavior. (Recall that attributing content to these mechanisms allows the theory to address these questions posed at the personal level.) Attribution-theoretical and developmental explanations, on the other hand, are pitched at the personal level through and through. It is not surprising that propositional attitudes are their primary explanatory coin. They are, in effect, scientific elaborations of folk psychology.

(2) Finally, recall Stich’s remark about folk psychology and the Autonomy Principle: “But if it is not clear whether the [autonomy] thesis is defensible, it is clear that if the thesis is accepted then . . . folk psychology is in trouble” (1996b: 23). He reasons as follows:

For . . . folk psychology includes lots of nomological generalizations that are couched in terms of the content of intentional states. But the Twin Earth argument (putatively) demonstrates that content does not supervene on the nonrelational physical properties of an organism. And the [autonomy] principle insists that the properties invoked in the generalizations of scientific psychology must supervene. So if scientific psychology has it right, then folk psychology must have it wrong. (1996b: 23)

Stich wonders whether the Autonomy Principle is best construed as a metaphysical principle or as a methodological principle, and then confesses that he is less than clear where methodology ends and metaphysics begins. I am sympathetic with the general point, but it is hard to make sense of the Autonomy Principle as anything other than methodological. There are many perfectly good properties that do not supervene on the current physical state of the organism, for example the property being a US citizen. The Autonomy Principle says that such properties should not be invoked in scientific psychology. But folk psychology is not scientific psychology. There is no conflict here, just two different taxonomies, answering to different explanatory concerns.

The Qualified Autonomy Principle (QAP) is explicitly methodological. It says that if a theory is concerned primarily to characterize the mechanisms underlying our cognitive capacities, then narrow taxonomies are preferable. Folk psychology is not concerned to characterize the mechanisms underlying our cognitive capacities. At least it is not primarily so concerned. It leaves that work for scientific psychology. QAP, then, simply doesn’t apply to folk psychology.
There is a second, and I think more interesting, response to Stich’s argument that folk psychology and the Autonomy Principle are in conflict. As Stich notes, folk psychology includes lots of nomological generalizations that are couched in terms of the content of intentional states. But while folk psychological generalizations are typically couched in terms of content, it is not obvious that the states themselves are individuated in terms of their content. Let me elaborate.

The fact that folk psychology identifies beliefs and desires by their contents does not imply that propositional attitudes have their contents essentially; in other words, it does not imply that content properties are individuative of the attitudes. To serve their typical predictive and explanatory functions, propositional attitudes must be construed as causally efficacious internal states of organisms, individuated essentially by the roles they play in mediating perception, cognition, and action. In other words, it is the functional roles of these internal states that are essential to their role in commonsense predictions and explanations of behavior. The generalizations of folk psychology provide only a partial and informal characterization of these complex functional roles. The most convenient (in practice, the only) way to refer to propositional attitude states is by their contents. The state that is typically caused by looking out the window on a rainy day and that typically causes (in conjunction with certain other functionally characterized states) umbrella-carrying behavior, can be referred to as the belief that it is raining. (I have mentioned only a small part of the complex functional role of this state, but when I identify the state by its content we can infer quite a bit about how it will interact with environmental conditions and other internal states to produce additional mental states and behavior.) The important point is that the contents of propositional attitudes play a reference-fixing role, enabling us to refer to internal states which future scientific psychology may eventually characterize in non-contentful terms, by elaborating precisely their functional roles.  

Contents, on this view, serve primarily to index mental states for the purposes of predicting and explaining behavior, but they are not essential properties of those states. It is possible for the same functional state – that is, for the same belief or desire – to be identified by reference to different propositional contents. For example, the type-identical belief states underlying my behavior and my Twin Earth counterpart’s when we both utter the form of words ‘water is wet’ are picked out, in our respective communities, by different broad contents. These different contents serve to pick out the same underlying functional state, much as different numbers – say 212 and 100 in the Fahrenheit and Celsius scales respectively – can be used to pick out the same underlying physical magnitude.

This construal of folk psychology obviously needs elaboration and defense, which I will not undertake here. But if the construal is correct, then folk psychology can respect the Autonomy Principle. Its taxonomic scheme can supervene on current physical states of the organism, even if the contents in terms of which its generalizations are couched do not.

Notes

1 See Dennett 1982 for a discussion of “east pole” and “west pole” positions.
2 Stich’s 1983 commitment to strong eliminativism is qualified. (2) and (4) are claimed to be “a serious possibility” (p. 242).
3 “A belief or memory storage system is modular to the extent that there is some more or less isolatable part of the system which plays (or would play) the central role in a typical causal history leading to the utterance of a sentence” (pp. 237–8). A similar premise— that folk psychology assumes that beliefs are stored in a modular fashion— plays a crucial role in Ramsey, Stich, and Garon’s 1991 argument for the claim that distributed connectionist cognitive models support eliminativism.

4 On holistic accounts of concept individuation, if the associated beliefs are not very similar, then the two subjects express different concepts by their use of the vocable ‘liberal’.

5 This formulation is from Stich 1991: 239. See also 1983: 164 and 1996b: 23.

6 Stich cites the case of Helen Keller, who comes to believe, after being told by a trusted informant, that there is a fat cat in the room. A sighted individual, upon seeing the cat, will acquire a belief with the same wide content and, hence, on one popular account of narrow content—the view that takes narrow contents to be functions from contexts to wide contents—a belief with the same narrow content. But the causal roles of these two belief tokens are clearly different: Helen Keller’s beliefs are never caused by visual states (1991: 246).

7 An adequate account must allow for misrepresentation, and for fine-grained determinate contents that distinguish rabbits, rabbit stages, and undetached rabbit parts.

8 In 1996b Stich canvasses various “semantic” arguments for eliminativism, including “the heterogeneity of the content taxonomy,” which is an amalgam of his own earlier arguments that content is unsuitable for use in science. He neither endorses nor rejects that argument.

9 See the visex/audex example discussed in the next section.

10 Hence the term “individualism,” coined by Burge 1979, is somewhat misleading as a name for the position defended here.

11 I am not suggesting that evolutionary biology and geology are not in the business of explaining how complex systems work, rather that their primary concern is to explain particular historical processes—those involved in the formation of terrestrial species and terrestrial geological structures, respectively. Given these explanatory goals, the increased generality that accrues to an explanatory scheme that relies on narrow taxonomies is not necessarily a virtue.

12 Segal 1989, 1991 characterizes a type of narrow content that he claims plays an individuative role in Marr’s theory of vision. Most interpreters of Marr’s theory, however, claim that Marrian contents are broad. (See, for example, Davies 1991, Egan 1991, and Shapiro 1993.) Segal 2000 characterizes a type of narrow content, which he claims is just ordinary content. It is, therefore, unlikely to be free of the $R^*$ properties. Chalmers 2002 characterizes a type of narrow content, which he calls epistemic content, but he makes no claim that epistemic content is actually used by cognitive science, and the claim is not independently plausible.

13 A full characterization of a computational mechanism also specifies the algorithm used to compute the function, and how the mechanism is physically realized in the brain or in a computer.

14 The claim that the canonical description of a computational device is not a semantic characterization needs an obvious qualification. Given that the canonical description specifies the mathematical function computed by the device, it 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 of a visual mechanism assigns visual contents to the states it characterizes. For example, it may interpret some structures as representing visible edges in the scene. A parsing theory will assign appropriate linguistic contents. It will interpret some structures as noun phrases, others as verb phrases. The canonical characterization prescinds from these contents.

15 This mechanism would not enhance the fitness of its possessor in the counterfactual environment. Of course, we assume that our own cognitive mechanisms are adaptations; we have
them because they enhanced fitness in the ancestral environment. But being an adaptation is a contingent property of any computationally characterized mechanism. Moreover, the Marrian mechanism, or any computationally characterized mechanism, is only contingently a visual mechanism. It enables its possessor to detect useful distal properties by transducing light in some, but not all, environments. In particular, it must do so for the actual environment. We shouldn’t be surprised, for example, that our visual system does not allow us to recover 3-D structure in an Ames room. The mechanism isn’t malfunctioning in this context – it works the same way it always does. But seeing requires a certain fit between mechanism and environment. In the normal case, of course, this fit is a product of evolution.

16 See Egan 1995a for argument.
17 Or, precisely because the functional roles of propositional attitudes are so complex, they may not be tractably specifiable. The view sketched here is agnostic about whether computational cognitive science will eventually provide a vindication of the propositional attitudes. Proponents of Fodor’s Representational Theory of Mind assume that it will, and that the vindication will take the form of a specification of the language of thought.
18 Readers will recognize here a “measurement-theoretic” construal of propositional attitudes, suggested by Churchland 1979, Dennett 1982, and Davidson 1989, among others. See Matthews (1994 and forthcoming) for elaboration and defense of the view.

References

Matthews, R. J. (forthcoming) *The Measure of Mind*.


