# 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.[1] Stich was right to reject eliminativism, though I am not sure that he rejected it for the right reasons. Stich 1983 contained a scathing 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 explicit. 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 him. I begin by sketching Stich's 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, even though Stich's original case against content remains largely intact. 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.[2] He also assumed then 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 subserves verbal and non-verbal behavior (p. 231), and that belief organization and storage is *modular*.[3] 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* that explains, together with the content-appropriate desire (to drink a beer, or perhaps just to drink something cold), 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 *observe-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*. A pair of beliefs is *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.[4] A pair of beliefs is *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 be taken to be expressing different beliefs when they say "granite counter-tops are durable."

Let us call these various properties of content ascription – 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.[5] 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.[6] Beliefs and desires violate the autonomy principle and are therefore not appropriate for use in a science of behavior. Hence, *weak eliminativism* is true – mature scientific psychology will not invoke beliefs and desires.

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 A theory that individuates states in terms of their content 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 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.[7]

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, would not make good science. And he still endorsed weak eliminativism – scientific psychology will not invoke commonsense intentional states because they 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 has 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.[8] 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. 1), 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 rather dubious assumptions about the reference of theoretical terms. But the failure of the inference undermines neither the 1983 or the 1991 anti-content arguments. It provides no reason to doubt that content has the *R*-properties, and 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 autonomy principle. In both 1991 and 1996b he holds the 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 – whether 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.[9]

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 (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 schemes that supervene on current physical states of the organism; therefore
3 a theory concerned primarily to characterize the mechanisms underlying our cognitive capacities should not individuate by 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). It has 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 appropriately 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 *modified autonomy principle* (henceforth, QAP) holds that a psychological theory concerned primarily to characterize the mechanisms underlying our cognitive capacities should employ taxonomic schemes that supervene on current physical states of the organism.

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 enjoined 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.[10] The resulting theories will be *radically internalist*; the boundary of the subject's skin will have no particular *individuative* significance.[11]

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.[12] 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 original 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. Doppelgangers 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.

Of course, it is only ordinary, broad content that 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.[13]

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; however, the content suited to play this role is *not* narrow (5 in my argument). It is not shared by all doppelgangers. It is able to play this explanatory role precisely because it is sensitive to features of the subject's context. I defend these two 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 may, 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 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, what the device does. 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 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]

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 sensitive to various aspects of the subject's environment, 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? The semantic interpretation is necessary to explain how the abstractly characterized process, in a certain (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, ultimately, 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 distal properties (*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*. The semantic interpretation forms a bridge between the intentionally characterized explananda of the theory and the abstract, mathematical characterization of the device 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 may 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. The content is (in part) externally determined; these structures cannot be expected to track, hence to represent, this property in every possible environment. In some weird counterfactual environments they may 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 bizarre accident or an experiment by IBM) it would not

enable its possessor to *see*. The mechanism would still compute the same abstractly characterized 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. It is an oft-noted fact about interpretation that it 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. This interest is itself reflected in the pre-theoretic explananda that define the theory's domain. Moreover, a given mechanism will subserve a cognitive function, such as vision, only in certain environments. The content ascribed to states of the mechanism can therefore be expected to be environment-specific, enabling an explanation of how the mechanism succeeds at recovering information about its normal environment. The semantic interpretation specifies which properties are tracked by states of the mechanism when it is functioning properly in its normal (internal and external) environment.

I hope the point is sufficiently clear for theories of perception. But it 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. But given that the explananda of psychological theories will be expressed 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, then we should expect the semantic interpretation that enables the theory to address these pre-theoretic explananda (which, of course, reflect our explanatory interests) to be specific to our world. Semantic interpretations appropriate to me and my twin-earth counterpart would assign different broad contents to our type-identical computational states.

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, will invoke 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 computationalism 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 *sub-doxastic* states.

The states involved in *domain-general* processes, such as the processes underlying decision making and rational revision of belief in response to new information, 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 slim.

In any event, nothing in *current* computational cognitive science 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 cognitive science, on 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. Moreover, folk psychology is concerned with much more than the theoretical goals of 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 simply no reason to think that these rich and complex notions will serve the rather austere explanatory purposes of computational cognitive science.

Of course, it is always possible to expand the list of propositional attitudes to include explanatory constructs from scientific psychology. While it is unlikely that the average person *believes* that objects are rigid in translation (Shimon Ullman's *rigidity assumption*, invoked in his 1979 explanation of the computation of 3-D structure from motion), and that acceptable grammatical transformations are governed by the minimal link condition (a principle of the minimalist program in syntax), we might, following Chomsky (1980), say that she (or her structure-from-motion mechanism and syntactic processor, respectively) *cognizes* such facts. But *calling* cognizing a propositional attitude doesn't make it one. Beliefs, desires, intentions, wishes, wants, 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 or convincing rationale.

The above considerations support the claim that computational cognitive science is unlikely to invoke (genuine) propositional attitudes. But *other* branches of scientific psychology *do* invoke beliefs and desires in their predictive and explanatory apparatus, 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 may 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 cognitive science, which I claim does not. Computational cognitive science, as noted above, is in the business of explaining how the mechanisms underlying our cognitive capacities work. The strategy for understanding how complex systems work is *functional decomposition*, which requires construing the behavior of the system as the outcome of independently specifiable components and operations. Thus, a computational theory may *begin* 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. Attribution-theoretical and developmental explanations, on the other hand, are pitched at the personal level through and through. To be sure, these theories are concerned with the characterizing the cognitive capacities of subjects, but they are not attempting to explain how the mechanisms underlying these capacities work[16] – they are not seeking mechanistic explanations – so functional decomposition is not apposite. In seeking personal level explanations of behavior, they are, in effect, scientific elaborations of folk psychology. It is hardly surprising, then, that propositional attitudes are their primary explanatory coin.

(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 lots of perfectly good properties that do not supervene on the current physical state of the organism, for example 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.[17] 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. We refer to 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 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.[18] 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'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 centigrade 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.[19] 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 I believe that *water* is wet. My twin who lives in a world where a different chemical compound fills the oceans and rivers has a distinct belief.

7 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).

8 An adequate account must allow for *mis*representation, and for fine-grained determinate contents that distinguish *rabbits*, *rabbit stages*, and *undetached rabbit parts*.

9 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 the argument.

10 The initial visual filter described by Marr 1982, and discussed briefly below, illustrates the point. The filter is characterized in mathematical terms: it computes the Laplacean convolved with the Gaussian; in effect, it computes a curve-smoothing function. The device is part of the visual system; however, we can speculate about whether the auditory system contains the *same type* of mechanism. The surrounding organism, including the rest of the visual system, is just more environment, as far as individuation is concerned.

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

12 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.

13 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.

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 Such a 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 only 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 So it would not be surprising if they do not respect the (qualified) autonomy principle.

17 See Egan 1995a for argument.

18 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.

19 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

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

Chalmers, D. J. (2002) *The Components of Content, in Philosophy of Mind: Classical and Contemporary Readings*. New York: Oxford University Press, pp. 608–33.

Churchland, P. M. (1979) *Scientific Realism and the Plasticity of Mind*. Cambridge, UK: Cambridge University Press.

Chomsky, N. (1980) *Rules and Representations*. New York: Columbia University Press.
Davidson, D. (1989) "What is Present to the Mind," in J. Brandl and W. Gombocz (eds.), *The Mind of Donald Davidson*, Amsterdam: Editions Rodopi, pp. 3–18.
Davies, M. (1991) "Individualism and Perceptual Content." *Mind* 100: 461–84.
Dennett, D. C. (1982) "Beyond Belief," in A. Woodfield (ed.), *Thought and Object: Essays on Intentionality*, Oxford: Clarendon Press, pp. 1–95.
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. (1995a) "Folk Psychology and Cognitive Architecture." *Philosophy of Science* 62, 179–96.
Egan, F. (1995b) "Content and Computation." *Philosophical Review* 104: 181–203.
Egan, F. (1999) "In Defence of Narrow Mindedness." *Mind and Language* 14(2): 177–94.
Fodor, J. A. (2000) *The Mind Doesn't Work That Way*. Cambridge, Mass.: MIT Press.
Heider, F. (1958) *The Psychology of Interpersonal Relations*. New York: Wiley.
Marr, D. (1982) *Vision*. New York: Freeman Press.
Matthews, R. J. (1994) "The Measure of Mind." *Mind* 103: 131–46.
Matthews, R. J. (forthcoming) *The Measure of Mind*.
Putnam, H. (1975) "The Meaning of 'Meaning,'" in K. Gunderson (ed.), *Language, Mind and Knowledge: Minnesota Studies in the Philosophy of Science*, Vol. 7, Minneapolis: University of Minnesota Press.
Ramsey, W., Stich, S. P., and Garon, J. (1991) "Connectionism, Eliminativism, and the Future of Folk Psychology," in W. Ramsey, S. P. Stich, and D. E. Rumelhart (eds.), *Philosophy and Connectionist Theory*, Hillsdale: Lawrence Erlbaum Associates, pp. 199–228. [Reprinted in Stich 1996a.]
Segal, G. (1989) "On Seeing What is Not There." *Philosophical Review* 98: 189–214.
Segal, G. (1991) "Defence of a Reasonable Individualism." *Mind* 100: 485–93.
Segal, G. (2000) *A Slim Book about Narrow Content*. Cambridge, Mass.: MIT Press.
Shapiro, L. (1993) "Content, Kinds, and Individualism in Marr's Theory of Vision." *Philosophical Review* 102: 489–513.
Spelke, E. (1990) Origins of Visual Knowledge, in D. Osherson and H. Lasnik, (eds.), *Visual Cognition and Action: An Invitation to Cognitive Science*, Vol. 2, Cambridge, Mass.: MIT Press.
Stich, S. P. (1983) *From Folk Psychology to Cognitive Science: The Case Against Belief*. Cambridge, Mass.: MIT Press.
Stich, S. P. (1991) "Narrow Content Meets Fat Syntax," in B. Loewer and G. Rey (eds.), *Meaning in Mind: Fodor and his Critics*, Oxford: Blackwell, pp. 229–54.
Stich, S. P. (1992) "What is a Theory of Mental Representation?" *Mind* 101: 243–61.
Stich, S. P. (1996a) *Deconstructing the Mind*. New York: Oxford University Press.
Stich, S. P. (1996b) "Deconstructing the Mind," in Stich 1996a: 3–90.
Stich, S. P. (1996c) "Naturalism, Positivism, and Pluralism," in Stich 1996a, pp. 192–200.
Stich, S. P. and Lawrence, S. (1994) "Intentionality and Naturalism," in French, Uehling, and Wettstein (eds.), *Midwest Studies in Philosophy* 19. [Reprinted in Stich 1996a.]
Stich, S. P. and Ravenscroft, I. (1994) "What is Folk Psychology?" *Cognition* 50: 447–68. [Reprinted in Stich 1996a, pp. 115–35. Page reference is to this edition.]
Ullman, S. (1979) *The Interpretation of Visual Motion*. Cambridge, Mass.: MIT Press.
Weiner, B. (1990) "Attribution in Personality Psychology," in L. A. Pervin, (ed.), *Handbook of Personality: Theory and Research*, New York: Guilford Press, pp. 465–85.