Mind & Language
Vol. 5 No. 3 Autumn 1990

ISSN 0268-1064
© Basil Blackwell

Article
*The Language of Thought: No Syntax Without Semantics**

TIM CRANE

Many philosophers think that being in an intentional state is a matter of being related to a sentence in a mental language—a 'Language of Thought' (see especially Fodor 1975, 1987 Appendix; Field 1978). According to this view—which I shall call 'the LT hypothesis'—when anyone has a belief or a desire or a hope with a certain content, they have a sentence of this language, with that content, 'written' in their heads. The claim is meant quite literally: the mental representations that make up the items of this language have semantic *and* syntactic properties. This is why, according to this view, cognitive psychology does, and should, treat that part of the mind which deals with intentional states as a semantic and a syntactic 'engine'.

Stephen Stich 1983 has argued at length that although there are good reasons for thinking of the mind as a syntactic engine, cognitive psychology should abandon the idea of a semantic engine. In particular, he argues that the 'folk' psychological notion of the semantic *content* of an intentional state should have no place in a mature science of the mind. He does think that there is a language of thought, but it is a purely syntactic language, a language without semantics. In this paper I will explain why he is wrong.

Patricia and Paul Churchland have also attacked the scientific credentials of folk psychology (see P.M. Churchland 1981; P.S. Churchland 1986). But unlike Stich, they do not think that the functionalist idea behind his

* I am very grateful to Matthew Buncombe, Jeremy Butterfield, Naomi Eilan, Tony Halbert, Roland Kenyon, Adrian Moore, Dennis Stampe, two anonymous reviewers for *Mind and Language*, and the audiences at the Cambridge University Moral Sciences Club and the Edinburgh University Centre for Cognitive Science for comments and discussion of this topic. Special thanks are due to Hugh Mellor and Gabriel Segal.

'syntactic theory of the mind' has any scientific credentials either. I think that the Churchlands are mistaken, but I will not argue against them here. Nor will I argue conclusively for the LT hypothesis. My aim is to offer an understanding of that hypothesis that removes some obvious objections to it, and to give a general account of the relation between the syntax and the semantics of the putative language of thought. A consequence of this account is that if there is a language of thought, then it must have semantics as well as syntax. So there cannot be a syntactic theory of the mind.

A number of philosophers have expressed scepticism about the idea that the syntax of intentional states, if they have any, can be completely independent of their semantics (see *e.g.* Dennett 1977, p. 103; 1982, p. 16; Blackburn 1984, p. 29; 1986, p. 121; Schier 1986, p. 275). In this paper I provide a general argument to buttress this scepticism. The general lesson of this argument is that there is no intermediate position between the Churchlands' eliminative materialism and content-based theories of the mind.

1. The Language of Thought

To begin with, we need to know what the Language of Thought is supposed to be, and why anyone should believe in it. Why should we apply the linguistic notions of syntax and semantics to the mind at all? We express our thoughts in spoken and written language, and we often talk to ourselves. But this obviously does not entail that thought has its own language. Nor does the fact that thoughts, like sentences, represent the world and thus have truth conditions and truth-values. For there are many non-linguistic ways in which aspects of the world may be represented: for instance, in pictures or in certain natural regularities (as the age of a tree can be represented by its rings). Why should we suppose that mental representation is linguistic?

The LT hypothesis starts from the idea that what makes thought like language is its compositional structure (see Fodor 1987, p. 138). It is uncontroversial that English, say, has such a structure: a finite number of primitive expressions are combined, according to certain rules, to form a potential infinity of complex expressions. The sentence 'Arthur loves Agnes' is made up of the constituents 'Arthur', 'loves' and 'Agnes', which also figure in the sentences 'Arthur eats cheese', 'Alf loves baseball' and 'Agnes walks to work'. The LT hypothesis claims that intentional states are compositional in precisely this way. It says that whenever an ascription of an intentional state is true, say 'Albert believes that Arthur loves Agnes', there is in the head of the thinker (Albert) a token representation corresponding to the content sentence ('Arthur loves Agnes'). This representation is composed of token mental representations corresponding to the words 'Arthur', 'loves' and 'Agnes', and other tokens of these types will occur in other

mental sentences when the thinker is in other intentional states whose contents concern Arthur, Agnes or loving.[1]

The LT hypothesis does not say that the mental sentence *is* the content of the intentional state in question. Rather, this sentence *has* a content, as natural language sentences do. The idea is that whenever anyone believes or desires or hopes (*etc.*) *that p*, a sentence which *means that p* is written in the head of the thinker.

Two questions immediately arise. First, what is it for a sentence to be written in someone's *head*? Any token sentence has a physical shape, the most familiar being the visible shapes of written sentences and the acoustic 'shapes' of spoken sentences. But, as modern computing vividly reminds us, sentences can be stored and transformed in patterns of electrical impulses—the possible 'shapes' of sentences are not restricted to these familiar ones. The LT hypothesis holds that sentences can be stored in the electrochemical states of the brain. In fact, the phenomena of modern computing provide more than an analogy to this aspect of the LT hypothesis: they provide part of its inspiration. Mental processes are, according to the hypothesis, actually *computational* processes. Thus when Albert believes that Arthur loves Agnes, the relations between this token intentional state and others in Albert's head are, according to the hypothesis, computational relations (see Fodor 1975, p. 198).

Second, what is it for a *sentence* to be written in someone's head? Someone might believe in mental representations and believe that computational relations hold between them, and still deny that these representations are sentences. But the LT hypothesis, for reasons I will give below, holds that mental representations must have *syntactic* structure as sentences in natural language do. The syntactic rules of the LT determine, among other things, which combinations of individual representations count as sentences of the LT.

The LT hypothesis does not give an account of what it is to *believe* that *p* rather than desire or hope that *p*. It is supposed to be a partial account of what it is to have an attitude with a particular content, not what it is to have a particular *attitude* to a content. Fodor's current view on this latter question is that to believe that *p* is to have a sentence that means that *p* written in one's 'belief box' (1987, Ch. 4; the 'box' metaphor is due to Schiffer 1987). What it is for a sentence to be written in one's belief box is for that sentence to be located in an appropriate place in that part of one's functional organisation that generates and processes beliefs. The idea is not entirely crystal clear, but I will say no more about it in this paper.

Why should we believe the LT hypothesis? Many philosophers find it utterly incredible. Dennett, for instance, has called it 'a gratuitous bit of

---

[1] The example is chosen for illustrative simplicity—I do not mean to imply that LT theorists think that LT sentences and their constituents correspond one-to-one with the sentences and words of the natural language the thinker may speak.

misplaced concreteness, a regrettable lapse into ideology' (Dennett 1978, p. 48). Although the LT hypothesis is supposed to be an empirical hypothesis, to be confirmed by psychological investigation, it is not mere speculation. Indeed, there are three good arguments in its favour, which I will briefly discuss.

The first is that it is this model of intentional states that cognitive science needs—and cognitive science has had great success and promises much. As Fodor has said, the 'intellectual breakthrough' of cognitive science 'needs the picture of the mind as a syntax-driven machine' (Fodor 1985, pp. 93–4). This is of course debatable; those who favour connectionist models of mental operations[2] may deny that cognitive science needs this picture, as Fodor himself acknowledges (1987, p. 139).

But—and this is the second argument—this conception of cognitive science has its roots in the claim that we must postulate a language of thought to explain how mental processes such as inference work; to explain, as Fodor nicely puts it, the 'similarities' between trains of thought and arguments (1985, p. 91). The best way to explain how the belief that *if p then q* plus the belief that *p* lead to the belief that *q*, is to think of the logical relation of entailment as being 'mirrored' by a causal relation between token sentences. The LT hypothesis claims that when someone has a sequence of beliefs of this form, what is going on in their head is a causal sequence of the LT sentences, 'if *p* then *q*', '*p*', and '*q*'. As Fodor says: 'mental processes are causal sequences of tokenings of mental representations' (1987, p. 17). These processes, Fodor claims, are governed by computational rules: algorithms that determine a representation as output, given a representation as input.

The third reason is that it is only by postulating a system of representations that we can explain the intentionality of thought from a physicalist standpoint. Thus Hartry Field argues that if we accept physicalism, and take the ascriptions of beliefs and desires to be literally true or false, then we must find some 'materialistically adequate' way of accounting for their truth-makers. If this cannot be done, then we are left with 'Brentano's problem': irreducibly intentional properties, which are a patent embarrassment to the physicalist. Field conjectures, therefore, that the truth-makers for ascriptions of (the explanatory causal roles[3] of) beliefs and desires are sentences in a system of internal representation, concretely realised in the physical structure of the organism. A token sentence is a concrete object, the sort of thing that can have causes and effects; it can play a role in the

---

[2] For an introduction to connectionism, see Clark 1989. For two good discussions of the relevance of connectionism to the LT hypothesis, see Cummins 1989 and Clark 1989a. The subject deserves more detailed treatment than I can offer here.

[3] For Field, a 'two-component' theorist, it is only the 'internal' component of an intentional state that is realised in the physical structure of the sentences in the LT (see 1978, pp. 79-88). For the distinction between 'explanatory causal roles' and semantic properties of intentional states, see McGinn 1982, pp. 207–16.

causal structure of the mind. This 'inner representation hypothesis' is, Field argues, 'the only known possibility for solving Brentano's problem' (1978, p. 27).

A similar point may be made without endorsing Field's physicalism. One aspect of the problem of intentionality is that of showing how intentional states can have the causes and effects we know they must have. These causes and effects must go via intrinsic properties, for reasons I cannot go into here.[4] What is needed therefore is a 'causal surrogate' for the content, an intrinsic property of a thinker that will produce the appropriate effects when that thinker is in a state with that content. Sentences in the LT may be seen as such causal surrogates: they interact in a way that mirrors the logical interactions of the contents. This is what Fodor means when he says that there must be 'a mechanism for the relation between organisms and propositions' (1981, p. 202).

Of course, the need for causal surrogates alone does not entail that they must be sentences, any more than the need for physical realisation does. But if we combine the need for causal surrogates with Fodor's point about mental processes, we have the basis of a plausible motivation for the LT hypothesis. For if mental processes involving intentional states require the contents of these states to have a compositional structure, then if these contents need causal surrogates, it is plausible to suppose that the causal surrogates must themselves have some sort of corresponding structure. The LT hypothesis, as I see it, is the claim that the best explanation for the compositionality of thought is the postulation of a system of syntactically structured causal surrogates.

2. *Syntax and Semantics for the Language of Thought*

So that's what the LT hypothesis claims, and those are, in brief, the main arguments for it. But in order for the hypothesis to be at all plausible, its proponents have to say how the supposed language can have the defining features of natural languages: syntax and semantics.[5] Semantic notions—meaning, truth, reference—are extremely controversial, but for the present discussion, I will assume that there are two features that a semantic theory must explain. First, consider how Tarski characterised semantics: 'a discipline which, speaking loosely, deals with certain relations between expressions of a language and the objects ... "referred to" by those

---

[4] For material relevant to this claim, see Shoemaker 1980, pp. 207–19; Mellor 1981, pp. 107–25; Lewis 1983, pp. 355–70. I argue for the claim in Crane 1989.

[5] Both 'syntax' and 'semantics' are terms that are used ambiguously: they can refer to the bunch of syntactic or semantic properties that a language has, or they can refer to the theories that study these properties. The ambiguity is harmless, and it will be obvious from the context in what sense I intend my use of these terms to be understood.

expressions' (1944, p. 345). Semantics should be concerned with relations between words and things—meaning is at least partly a matter of reference. Second, it has been generally recognised since Frege that the meanings of sentences depend on the meanings of their parts, and that a theory of meaning should reveal this. These two points—that meaning is at least a matter of reference, and that meaning is compositional—should be generally accepted.

But how can the mind have semantics? There seem to be some crucial characteristics of the meanings of expressions in natural languages which could not apply to symbols in the brain. Consider, for instance, the much discussed claim that the attachment of meaning to our words is *arbitrary*. Since the semantic relations between words and things are not, on the face of it, fixed by any natural relations between them (such as resemblance), it is often proposed that the only way these relations can be fixed is by *conventions* holding among users of the language.

But how do the relatively arbitrary shapes that serve as words of a natural language acquire their conventional meaning? We must content ourselves here with truisms. If we agree that conventions are largely a matter of shared beliefs, it should be plain that it is the belief of language-users that determine how words are used, and thus what they mean. This is one uncontroversial understanding of the slogan 'Meaning is Use'. One could go on to say the meaning of a word is *constituted* by the use to which it is put (Wittgenstein may have thought this: *cf.* 1953, §432). Or one could say that it is the *product* of the intentional states of the user of the word (Grice 1957, Bennett 1976). But either way, the point is clear: sentences and words have meaning only because they are intentionally used by people. But such use cannot be what makes the symbols of the LT meaningful, since we surely do not intentionally use or 'manipulate' these symbols. And since we do not intentionally use them, the LT symbols cannot be arbitrary, and thus conventional, in the way natural language sentences are. LT sentences could only have their meaning in the way sentences of a natural language do if there were a regress of homunculi, manipulating symbols *ad infinitum*. And this is obviously unacceptable.

The regress can be blocked by insisting that unlike natural languages, sentences in the LT do not have to be *understood*. As Fodor puts it:

> By assumption [the LT's] formulas—unlike those of German—can have their effects on the course of thoughts *without having to be understood* . . . This is the trick which computational psychologists use to get the goodness out of postulating a language of thought without having the traditional problem of a regress of languages and interpreters. (1987a, p. 67)

Where languages are concerned, what you can't understand, you can't learn. So if the LT is not understood, then it is not learned either. One of Fodor's polemical slogans in defence of the LT is that you can't learn a

language unless you already know a language. But this doesn't mean that you can't know a language without *having learned* a language—rather, it means that the LT is *innate*. So two crucial differences between the LT and a natural language are that the LT is not understood, and that it is innate.

This response eliminates the problem of regresses. But now Fodor has to say how sentences in the LT can have semantics, if they are not had by being intentionally used. This is, as he acknowledges, the hardest task for the LT theorist. 'Of the semanticity of mental representations', he laments, 'we have, as things stand, no adequate account' (1985, p. 96). But Fodor can take comfort from the fact that the problem of meaning is a problem for everyone; perhaps there is no special problem for the LT. In particular, the two indisputable facts about semantics mentioned above—it is concerned with reference, and it must reveal language's compositional structure—seem to have a clear application to the language of thought. The compositionality of thought, as we saw, is one of the defining ideas of the LT hypothesis. As we saw, thoughts do seem to have constituents—the thought that Alf is hairy has elements in common with both the thought that Alf is corpulent and the thought that Harold is hairy: the constituents 'Alf' and 'is hairy'. And these constituents stand in relations of reference and predication to items in the world—Alf and hairy things respectively. How it is that symbols in the brain come to stand in these relations is indeed a difficult question. Fodor (1987, Ch. 4) has made an attempt to apply some of the ideas of causal theorists of meaning (especially Stampe 1977 and Dretske 1981) to answer it. Whether or not he is ultimately successful, it seems clear that there is not all that much wrong with the very idea of a semantics for the language of thought, once the problem of regresses is explained away.

So much then, for the general notion of semantics and how it might be extrapolated to the mind. What about syntax? It is natural to think that syntax has to do with the *form* of linguistic items, while semantics has to do with their *content*.[6] The notions of syntax and form are often used by participants in the mental representation debate, but rarely explicitly defined. In 'Methodological Solipsism', Fodor attempts to define syntax thus:

> What makes syntactic operations a species of formal operations is that being syntactic is a way of *not* being semantic. Formal operations are the ones that are specified *without reference to such*

[6] The association of 'formal' with 'syntactic' has had the confusing result that many philosophers (*e.g.* Carnap) have called 'syntactic' an enquiry that deals with formal properties such as consistency, logical consequence, *etc.*—properties that are best defined in semantic terms: *e.g.* in terms of truth-preservation. As Higginbotham says, 'The word "semantics" has acquired a restricted usage, one that would classify any notion definable by conditions on form as syntactic rather than semantic.' But, he adds, notions like consistency 'can be seen as syntactic only in a sense of syntax that takes in far more than the characterisation of well-formedness' (1988, p. 228). I entirely agree with Higginbotham's perceptive diagnosis of this issue, and I should point out

> *semantic properties of representations as, for example, truth, reference and meaning.* Since we don't know how to complete this list (since, that is, we don't know what semantic properties there are), I see no responsible way of saying what, in general, formality amounts to. The notion of formality will thus have to remain intuitive and metaphoric, at least for present purposes: formal operations apply in terms of the, as it were, *shapes of the objects in their domains.* (Fodor 1980, p. 227; my italics)

But it will not do merely to call syntactic properties of representations those which are identified without reference to their semantic properties. Any token representation will have many properties which can be identified without reference to its semantic properties, but these may not be syntactic properties (for instance, the representation's chemical properties). There are many ways of not being semantic which are not ways of being syntactic.

So what is syntax? In *Syntactic Structures*, Chomsky says that it is the 'study of the principles and processes by which sentences are constructed in particular languages' (1957, p. 11). The basic idea of a syntactic theory is that of a *well-formed* or *grammatical* expression. The elementary well-formed expressions, the constituents, combine to form well-formed sentences, and syntactic rules tell us which combinations of these constituents are grammatical and which are not. The task of the syntactic theorist is to say, for a given language, what its basic syntactic categories are, and which rules produce grammatical expressions from the systematic combination of members of these categories. The theory is thus capable of assigning a structure to every sentence of the language.[7]

In what sense will the LT have syntax? There are obvious uncontroversial links with this idea of syntax. Syntactic rules will determine whether any string of symbols in the LT is well-formed and which combinations of strings yield further well-formed strings. So far so good.

But why does Fodor say that syntax has anything to do with the *shapes* of linguistic items? In English, two sentences can have the same syntax and different shapes—just consider the difference between 'The pope likes fish' and 'THE POPE LIKES FISH'. So shape doesn't supervene on syntax. But does syntax supervene on shape? The syntax of English sentences certainly doesn't, as syntactic ambiguities show: 'Flying planes can be dangerous' has two correct syntactic parsings that can correspond to the same surface shape.

However, with thought the case is different, since there is, I take it, no such thing as an ambiguous thought. Thoughts cannot be ambiguous,

---

that I am not discussing this more liberal sense of syntax.

[7] What I say here about syntax is necessarily brief, but it serves my purposes. For a detailed introduction, see Sells 1985.

since a *sentence's* being ambiguous is a matter of its having two or more *thoughts* associated with it. To disambiguate a sentence is to say what these thoughts are. So if there is a *language* of thought, its sentences could not be ambiguous since that would require there to be a 'lower' level of thought to provide the disambiguations. So the phenomenon of 'same shape/different syntax' will not, it seems, occur in thought. Where a given LT is concerned, then, syntax will supervene on shape: there will be no difference in syntax without a difference in shape.

But why should shape matter to syntax at all? The connection is crucial to understanding the nature of the LT. In the previous section I claimed that one motivation for the LT hypothesis is the need to explain how intentional states can have the right sorts of causes and effects. Relations between sentences in the LT provide the causal mechanisms for mental processes such as inference. Moreover, immediate causation must go via intrinsic properties of objects. So the sentences in the LT must have their causes and effects in virtue of their intrinsic properties—Fodor calls these intrinsic properties of the sentences their 'shapes'. This is why he equates the notion of 'syntactic encoding' with the notion of 'representation by intrinsic features' (1987a, p. 80; *cf.* p. 67). The idea is that since it is in virtue of some intrinsic property—its 'shape'—that a symbol has effects, and since there is no difference in syntax without a difference in shape, then the syntax of the LT is actually the causal mechanism for the interactions of intentional states (see Pylyshyn 1984, p. 39). It is this use of the idea of syntax that makes the LT hypothesis an interesting one. Of course, Fodor is right to say that the idea of a shape of a symbol in the LT is somewhat metaphorical. But there is no reason to suppose that this metaphor cannot eventually be explained in non-metaphorical terms.

3. The Syntactic Theory of the Mind

It is this idea of the syntax of the LT as the mechanism of thought that inspires Stich's 1983 views. Stich urges that we should retain the picture of the mind as a syntactic engine, but abandon the notion of the content of an intentional state. He argues that difficulties in the individuation of content demands that we abandon folk psychology as a serious prototype for cognitive theories. The chief difficulty with the notion of content, according to Stich, is that ascriptions of content are crucially sensitive to their context, in such a way that strict scientific generalisations (which must abstract from context) are impossible to formulate. In the spirit of Quine, Stich argues that there is no strict standard of sameness of content at work in folk psychology; content-ascriptions rely rather on certain sorts of similarity that are too vague or context-based to be useful in science. His alternative theory would attempt to respect certain elements in the outdated folk psychology—for instance, certain patterns of (apparently) logical relations between the (apparent) contents of states. The difference

is that these elements would be more rigorously characterised in non-semantic terms. It is this non-semantic characterisation which Stich calls the 'syntactic theory of the mind' (STM).

Stich agrees with much that the LT hypothesis claims. He says that 'Fodor's basic idea is on the right track' (1983, p. 43) and two chapters of his book are devoted to defending the 'mental sentence' view of intentional states. The crucial difference is that the STM's language is 'no more than an infinite class of complex syntactic objects. *It has no semantics*' (1983 p. 153).

The STM is not itself a cognitive theory; it is a view about what cognitive theories are and ought to be like. The essential idea of the STM is that 'causal relations among cognitive states mirror formal relations among syntactic objects' in such a way that 'it will be natural to view cognitive state tokens as tokens of abstract syntactic objects' (Stich 1983, p. 149). The theory assumes that these tokens are in fact token brain states. The task for the STM theorist is threefold. First, the theorist must specify a class of syntactic object-types, in a way which assigns structure to these objects (p. 150). Secondly, the theory hypothesises that for each organism covered by the theory, there is a set of state-types whose tokens are 'causally implicated in the production of behaviour' (p. 151). These types are then mapped onto the syntactic object-types. Thirdly, the theorist spells out generalisations, detailing relations between hypothesised neurological states, indirectly by specifying 'formal relations' (p. 151) among the abstract syntactic object-types.

So far, much of what Stich claims is in the spirit of functionalism, as he readily acknowledges (p. 151). The distinctive feature of the STM, though, is that the indirect specification of causal relations makes no use of the notion of content or any other semantic notions. The generalisations made at this intermediate level—between content and neurophysiology—will appeal only to the purely formal properties of representations. This is why the theory is syntactic.

I think Stich's arguments against content are inconclusive. But I will not attack them directly. Instead I will attack Stich's assumption of what I shall call the *conceptual independence* of the syntax of the LT from its semantics: the thesis that it is possible for the LT to have syntactic structure without having any semantic interpretation. But before arguing against this, it is necessary to distinguish Stich's thesis from two more plausible, but irrelevant, theses.

The first thesis is that in a quite trivial sense, it is possible to have syntax without semantics. That is, you can construct a system of mere shapes organised by rules that say (a) which are the primitive shapes, and (b) which combinations of these primitives can be combined to make 'well-formed' complex shapes. And you can call these rules the 'syntax' of the system if you wish. But this shows nothing about whether the LT's syntax is independent of its semantics, for two reasons. First, it is plausible that we only make sense of such a system having syntax by thinking of it

as if it were a language, as if the shapes were *symbols*, and therefore as if it had semantics. We make sense of such a case derivatively, by analogy with languages that have semantics. Secondly, until the system is put to use in some way, the idea of its formulae being well- or ill-formed is entirely without substance. Being well-formed here is just a matter of stipulative definition. This is not so with real syntax, where what is or is not well-formed is discovered by hard work. And the same will apply to the syntax of the LT, if it exists.

What isn't trivial, as we shall see, is the claim that we can actually *discover* that the syntax of the putative LT is like this—that we can find out that it has a syntax but no semantics. This claim needs careful examination, and it cannot be proved simply by showing that it is *possible* for someone, in an idle hour, to construct a 'syntax' in the way described.

The second irrelevant thesis is Chomsky's thesis of the 'autonomy of syntax'. This is, in effect, the claim that syntactic theory should see the mechanism that generates the formal structure of language as an autonomous 'mental organ' with a somewhat modular structure.[8] Chomsky describes the thesis thus:

> [the] rules of syntax or phonology . . . are organised in terms of 'autonomous' principles of mental computation and do not reflect in any simple way the properties of the phonetic or semantic 'substance' or contingencies of language use. (Chomsky 1980, p. 246)

Whether this is true depends on the extent to which theories of syntax have to employ semantic considerations in formulating the correct syntax for natural language. But this has no bearing on Stich's claim that the LT has a syntax but no semantics. For to treat the mechanism of syntax as a modular mental organ in no way entails that there are not semantic interpretations too. And of course, all linguists, no matter what their view on the autonomy of syntax, think that natural language has semantics as well as syntax. This does not mean, of course, that the semantic theory for a natural language must employ all and only the semantic notions of folk psychology. But it will have to employ some, for the notion of *meaning* is common to folk psychology and to scientific semantics.

The autonomy of syntax is the straightforward empirical thesis that we get more of an insight into the workings of natural language by treating syntax as autonomous. But it is incontestable that the general aim is to understand language in all its aspects—syntax, semantics, pragmatics, phonology. As Chomsky has said, 'we will have only a partial understanding of syntax if we do not consider its role in the expression of thought,

---

[8] See Chomsky 1980, pp. 40–46, for the claim that the mechanisms of syntax are modular. For the general idea of a modular psychological structure, see Fodor 1983.

and other uses of language' (1976, p. 59). If syntactic mechanisms are autonomous, then it may be possible to describe the syntax of a language independently of its semantics. In this case, we should say that syntax and semantics will have *descriptive* independence, but not *conceptual* independence—since it is taken for granted that the language actually has a semantics. So if the thesis of the autonomy of syntax were true, it would not lend any support to Stich's thesis that an actual language (the LT) could have syntax without semantics.

4. Syntax Without Semantics?

So how can Stich establish the conceptual independence of the syntax and the semantics of the LT? It cannot be done by appealing to the trivial thesis that it is *possible* to call an arbitrary selection of rules governing some arbitrary selection of shapes 'syntactic'. What has to be shown is that states of mind have a causal structure that is syntactic in a non-trivial sense, and they have this structure without having semantic properties. In this section I will argue that this cannot be done.

To begin with, we should return to the question of how any 'mental sentence' theory (Fodor's or Stich's) will pick out the syntactic properties of its language. In Section 2 we saw that LT theorists characterise syntactic properties as those properties of a symbol that are causally relevant to inference and the production of behaviour—hence Fodor's equation of 'syntactic' with 'intrinsic'. But as we saw, these token representations will have many other intrinsic properties that are not thus relevant. If Fodor is right about the LT, then the syntactic properties whose tokenings cause behaviour will be intrinsic properties, but this doesn't mean that any intrinsic property is syntactic. (To put the point in terms of the metaphor of *shape*: since not all aspects of the shape of a symbol are aspects of its syntax, the idea of shape alone does not pick out syntax.)

How, then, can a theory establish which intrinsic properties are syntactic? Fodor can answer this along the following lines. Take an intentional state of an agent, say Alf's belief that *eggs are poisonous*. What makes Alf's state a *belief* that eggs are poisonous is that an LT sentence with that meaning is written in his 'belief box'. What makes this sentence mean that eggs are poisonous is, as we saw, a hard question, but one we may ignore. The idea is that Alf has a sentence 'eggs are poisonous' written in his head, and this sentence has a semantic and syntactic structure. Its semantic structure consists in the fact that tokens of each of its type of constituent can occur in the contents of Alf's other intentional states—for instance, in his hope that the eggs he ate last week won't poison him. Its syntactic structure consists in the fact that all his beliefs (and desires *etc.*) concerning eggs and poison are *causally* related in a systematic way. For instance, Alf's belief that *this is an egg*, plus his belief that *eggs are poisonous*, plus his desire *not to eat poisonous things*, together cause him to form the

intention *not to eat this*. The combination of these beliefs and this desire cause the intention because their common semantic constituents (Alf's 'concepts' of egg and poison) correspond to common intrinsic properties of symbols in Alf's LT—their shapes.[9]

On Fodor's view, then, the syntactic properties of Alf's intentional states in this case are those intrinsic properties of Alf that causally interact to produce his egg-avoiding intention and behaviour. Of course, there are many of Alf's intrinsic properties that are involved in the causation of this behaviour—for instance, properties of his motor centres and muscles—but what the LT hypothesis aims to explain is the causation of behaviour by *intentional* states. So although it is possible that a similar body movement to Alf's egg-avoiding one can be produced not by beliefs and desires but by some other constellation of causes, this does not show that *these* causes are syntactic properties, since the behaviour is *ex hypothesi* not intentional. The syntactic properties are the properties of Alf that are causally sufficient and necessary in the circumstances for his behaviour to be a piece of intentional egg-avoiding behaviour.

Fodor's theory can thus specify which intrinsic properties of symbols in the LT are their syntactic properties: they are those properties of the symbols in virtue of which those symbols play a causal role in inference and cause intentional behaviour. Syntax, on this view, explains how the semantics of the symbols in the LT is relevant to the explananda of a theory of intentional states: principally, to inference and behaviour. Syntax does not *replace* semantics; it *vindicates* it.

Saying this much about the relation between the syntax and semantics of the LT does not mean that a theory of the mind that postulates an LT has to employ all and only the semantic concepts of folk psychology. It does not have to employ them all because there may be some concepts used in folk psychology that have little application in its scientific descendant. And it is unreasonable to suppose that an LT theory has to employ only folk psychological concepts. Theories typically need to construct new concepts to give ever more sophisticated accounts of the phenomena; the theory of mind should be no exception. But the LT hypothesis follows folk psychology in one crucial feature: it treats intentional states as having content.

Nor does the LT hypothesis' conception of the relation between syntax and semantics entail that *all* syntactic properties of representations can be explained in semantic terms. This is not so in natural language—indeed, it is part of what the thesis of the autonomy of syntax means. So since some representations in natural languages have syntactic properties that

---

[9] I say *Alf's* LT because it may be the case that in different people, different shapes correspond to the same syntactic property—syntax may be 'variably realised' by shape, from person to person, and thus LTs may have to be relativised to a person. The LT hypothesis is in too much of an embryonic form to be able to settle this question at present.

cannot be explained in semantic terms, it should not be a surprise if the LT has some too. But this in no way undermines the Fodorean picture as I have presented it.

So how can Stich answer the question 'Which intrinsic properties are syntactic?', if he cannot appeal to semantic properties, as Fodor can? As I see it, there are only two possible general strategies for the STM. The first, which I shall call the 'moderate strategy', is to adopt the approach to this question I attributed to Fodor, and then argue that semantics is dispensable on this approach. The second, which I shall call the 'radical strategy', is to adopt an approach entirely different from Fodor's. I will argue that if one takes the first strategy, one cannot avoid semantics, and if one takes the second, the idea that the mind has syntactic structure becomes entirely vacuous.

We can examine the prospects of the moderate strategy for the STM by looking at Stich's example of how an STM will replace belief-desire generalisations. (To bring out the parallel with Fodor's theory, I will use a similar example to the one I used above.) Suppose this time that (in folk psychological terms) Alf's belief that *eggs are wholesome* and that *this is an egg* plus his desire *to eat wholesome things* cause him to eat this egg. Stich's example of an STM says that Alf is in a number of 'B-states' ('roughly analogous to beliefs') and a number of 'D-states' ('roughly analogous to desires'—Stich 1983, p. 154). In particular, he is in a B-state (B1) that is 'mapped onto' a syntactically well-formed formula (wff) that 'roughly corresponds' to the natural language sentence 'eggs are wholesome'. And he is in a B-state (B2) that is mapped onto a wff that corresponds to the natural language sentence 'this is an egg'. He is also in a D-state (D1) that is mapped onto a wff roughly corresponding to the English sentence, 'I shall eat wholesome things' or some such. Alf's being in states B1, B2 and D1 cause him to eat his egg. Moreover, this particular explanation of his action is subsumed under true generalisations relating B- and D-states (1983, p. 155).

What Stich offers here is a sketchy example of how states of mind are supposed to have syntactic structure alone: the wffs are 'built from a number of primitives by a limited number of compounding operations' (Stich 1983, p. 154). But they do not, Stich says, have a compositional semantics. They are merely syntactically well-formed strings of uninterpreted shapes. So, in my example, although the perception of a particular egg causes Alf to be in B2, and the wff caused to be written in Alf's B-box has a constituent also tokened in the wff of B1 (the constituent corresponding to 'egg'), B2's wff does not *mean* that *this is an egg*, and the constituent does not mean *egg*.

This might seem incredible. After all, we are only making sense of this example by supposing that the B- and D-states 'correspond' to particular beliefs and desires. We are then supposed to ignore this correspondence and view the states as uninterpreted. I do not deny that we can do this; but does this act of imagination show that the B- and D-states really have

no semantic interpretation? As Higginbotham points out, we do not show the superfluity of semantics by 'disinterpreting a creature's language [and] observing that the creature will behave as we do (in the relevant respects)' (1988, p. 230).

It is not enough to say that we can take a case like this and describe the causal relations between the relevant states *without mentioning* their semantic properties. For Fodor will agree:

> it's got to be possible to tell the whole story about mental causation (the whole story about the implementation of the generalisations that belief/desire psychologies articulate) without referring to the intentional [*i.e.* semantic] properties of the mental states that such generalisations subsume. (Fodor 1987, p. 139)

So the mere fact that we can invent the idea of a B-state (*etc.*) and redescribe our example in terms of it does not show anything more than the *descriptive* independence of the syntax and the semantics of the LT. We have not shown that the STM can answer the question about which properties are syntactic without using semantic notions.

Stich will respond that the idea of wffs 'corresponding' to English sentences is a mere expository convenience. The claims about what the B- and D-states and their wffs correspond to 'are no part of the STM explanation' (1983, p. 175). They are intended only to show how STM explanations can parallel folk psychological ones, but they are not essential to the STM. Although we might use semantic notions in answering the question 'which intrinsic properties are syntactic?', these notions do not characterise anything of substance. They are mere heuristic devices.

So Stich might say: 'these properties are syntactic because they have a certain structure; and items with this sort of structure are those items that Fodor, influenced by the naive folk theory, claims are LT sentences with content. But as a matter of fact, we know better: the beliefs of folk theory and their contents cannot be used by a serious science of the mind. However, since the states picked out by the folk theory seem to be relevant to the causation of behaviour, and a theory of mind must explain behaviour, we should use folk theory as a rough guide to the individuation of our syntactic properties. The appeal to folk psychology is thus a useful (though ultimately incoherent) guide through a field where little has been discovered. We need this guide because we need to succeed where folk psychology has failed—in fully explaining behaviour. In the future, however, we expect that this particular ladder can be thrown away' (*cf.* Stich 1983, p. 183).

But this raises two problems about what is to be explained by the STM, the second more important than the first. First, all parties in the dispute must accept that one thing a theory of mind must explain is *linguistic* behaviour. Much of our intentional traffic with the world is linguistic: predominantly in communication, but also in phenomena such as thinking

in words. Among the many things thinkers do with a language is to use its sentences to express what are, on the face of it, the contents of their beliefs and other intentional states. So how can the STM explain linguistic behaviour, if it leaves no place for the idea of content in any explanation of the mind?

There are two answers the STM can give. The first is that just as semantic notions are disreputable in the theory of mind, so they are disreputable in the theory of language: we should have a syntactic theory of language as well as a syntactic theory of mind. The second is that language is beyond repair: there can be no science of language, since it involves notions that are just too imprecise to be acceptable in any science.

Both these answers are really very implausible. The thesis that linguistics consists just of syntax (and presumably phonetics) would entail that all linguists are radically mistaken about what they are studying. To show this would take more argument than Stich has given. He does not commit himself explicitly on this issue in his book, and his only remark is rather tentative:

> although Fodor and Pylyshyn are surely right that much linguistic behaviour can be described in intentional or semantic terms, it is far from clear that cognitive psychology ought to aim at explaining this behaviour *described in this way*. (Stich 1983, p. 195)

This is not very satisfactory unless Stich can say exactly how cognitive psychology should describe linguistic behaviour, which he does not. The fact that much linguistic behaviour can be, and is, described in 'semantic or intentional terms'—*i.e.* in terms of meaning—is a *datum* that needs to be explained. This is why there is a branch of linguistics called 'Semantics'—the component of a theory of language that classifies types of sentences and words in terms of their meaning, and then postulates mechanisms to explain how meaning works.

Again, it should be emphasised that this does not mean that the semantic theory for a natural language has to employ only the semantic concepts of folk psychology. Semantics, as a component of linguistic theory, will use many concepts that have no use in folk psychology, or no straightforward folk psychological interpretation. But there must be some overlap, since the notion of *meaning* or *content* is shared by both theories. Without this overlap, using the term 'semantics' in both the folk psychological theory of mind and the theory of language would merely be a pun. But as we saw in Section 2, it is not. All agree that folk psychology, whatever its merits, employs the idea of the content of an attitude, and that it holds that thinkers use sentences to express these contents. And semantics is—by definition—an attempt to explain the meanings of words and sentences. To do so will of course involve many technical concepts that are not part of folk psychology. But there should be no doubt that there is some similarity in the explananda of semantics and folk psychology—namely

that, on the face of it, sentences, like contents, are about things in the world, and have truth-values. Stich will of course say that to explain this apparent fact, we do not have to invoke intentional states with contents. But then he has to explain away the apparent fact, which he has not done.

So it is more likely that Stich will say that there can be no science of language, a view not without adherents.[10] But if this is so, then he is not entitled, as Fodor is, to assume that B-states and D-states have a *syntactic* structure like the syntactic structure of (say) English sentences. Fodor is entitled to conjecture that the LT may have a structure like the structure of a natural language, and its syntax may well resemble natural language syntax to some extent—the lexicon may contain LT analogues of noun phrases, verb phrases, determiners *etc.* This may not be true, but it is a reasonable assumption for Fodor to make. But if Stich thinks, unlike Fodor, that natural languages cannot be systematically studied, then he should not assume without further argument that the wffs in the B- and D-boxes have syntactic structure that resembles the syntax of natural language sentences. So if he takes this view, he should abandon the moderate strategy and turn to what I call the radical strategy. This will be discussed below.

The second, more important, problem with the response I attributed to Stich has to do with non-linguistic behaviour. The moderate strategy for the STM is supposed to show how an account in terms of the 'syntactic analogues' of beliefs and desires, B-states and D-states, can explain behaviour as adequately as Fodor's theory does. And since Fodor's theory employs the idea of content, which the STM holds to be disreputable for independent reasons, the STM account is preferable (*cf.* Stich 1983, p. 158).

But the behaviour to be explained in our example—eating an egg—is described in intentional terms. This behaviour is, on any story, different from what Alf does when, say, under hallucination he picks up his spoon, cracks his egg and puts some in his mouth—even if the body movements in question are the same. It is no good for Stich to say that the cases are distinguished by the fact that in the first they are caused by the B-states and D-states in question, since we are supposed to find out what the B- and D-states *are* by seeing what they cause. Of course, there are many ways of describing a piece of behaviour. What Stich needs is a notion of behaviour that counts as type-identical all the body movements that Fodor's theory counts as (behaviourally) type-identical, and yet does not do this by classifying them in terms of the contents of the intentional states that cause them.[11]

Stich admits that 'substantial work needs to be done in forging appropri-

---

[10] This view comes in many forms. For a recent defence, see Schiffer 1987. One of Stich's biggest influences is Quine, whose thesis of the indeterminacy of translation (1960, Ch. 2) clearly entails that there can be no science of language.

[11] *Cf.* Higginbotham: 'from disinterpreted states we would expect to get only explanations of disinterpreted behaviour' (1988, p. 230).

ate autonomous behavioural descriptions for use in scientific psychology' (p. 169). But, except for saying that these will not be 'purely physical descriptions of movements', he does not enlighten us as to the character of these descriptions.[12] The only constraint seems to be that the descriptions should be true of a subject and his Twin-Earth Twin or 'replica' (pp. 168–70).

But the problem with this notion of behaviour is not just that Stich hasn't said what it is. The problem is that folk psychology classifies many pieces of behaviour as of the same type while allowing that many (if not most) of the other non-intentional properties of the behaving organism are not of the same type. It is a familiar fact that there are many ways I can greet a friend: by waving, saying 'hi!', punching him playfully on the arm, pretending to ignore him. I need have little non-intentional in common in all these cases for them to be tokens of the same type of action. And even with 'basic' actions, such as raising my arm, there need not be one open sentence couched in some non-intentional vocabulary true of all and only those actions which are arm-raisings.

The point is not new, and it or one like it has been made by Pylyshyn (1984, p. 161), Fodor (1981, pp. 25–6) and others in an argument for content. What they take it to show—and I think they are right—is that the psychologically interesting thing these actions have in common is the fact that they are caused by states with some of the same contents. This does not conflict with Fodor's 1980 formality condition, the claim that semantics supervenes on syntax. It is true that punching my friend on the arm might in one context be a playful greeting, while in another it will be an expression of irritation. But all the formality condition requires is that the behaviour should supervene on my total non-intentional state in that context. According to Fodor, this will include the formal properties of sentences in my LT. Since these sentences differ in content only if they differ in form, and since according to Fodor, the sentences in our two examples *will* differ in content—in one case, the sentence will have the content 'it's nice to see Jim again', in another, 'Jim is so irritating'—then they will differ in form. So my total intentional state will supervene on my total non-intentional state. One can accept the formality condition, and thus accept that behavioural descriptions are autonomous in Stich's sense, while still holding that content has an essential role in individuating behaviour.[13]

---

[12] 'I am not at all sure just what an appropriate autonomous description of [a] behavioural event would be. Indeed, perhaps an appropriate autonomous behavioural descriptive language does not yet exist. But no-one said psychology was supposed to be *easy*'. (Stich 1983, p. 196)

[13] Stich claims that Fodor cannot preserve his allegiances both to content *and* to the formality condition, for essentially the reason that Twin Earth cases show how content and form can come apart (1983, p. 197). My view, however, is that the Putnam-Burge Twin Earth arguments are not persuasive. One can keep the formality condition without having to opt for the dubious notion of 'narrow content'. See Crane 1989, Ch. 1.

So it is hard to see how the STM can count very different body movements as the same behaviour—and thus preserve the generalisations of folk psychology that it is supposed to—without seeing some of these body movements as the products of the states with the same content. Of course, since the formality condition says that there is no difference in content without a difference in form, then if Stich can find the formal or syntactic properties of his wffs, he can formulate all the generalisations that Fodor can without describing the contents of the wffs. But the question I am pressing is whether Stich can begin to find his B- and D-states without assuming that the behaviour they cause is intentional: that is, without assuming that what makes one piece of behaviour the same as another is the fact that their causes share content or elements of content. The point is not just epistemological—not just 'how will Stich *find out* which are the syntactic properties of the wffs without assuming that behaviour is type-identified by content?' Rather, the point is: 'what licenses Stich to think there are syntactic wffs *at all*, if he is not willing to accept that what individuates his explananda is something semantic?'

I agree with Stich that whenever one has a particular belief/desire explanation, one can recast it in terms of a B-state/D-state one. But as I said above, the mere availability of the B-state strategy does not by itself show the superfluity of content. It only shows that syntax and semantics have descriptive independence. Stich needs a way of picking out the B- and D-states that does not appeal to their contents. I have just argued that he cannot use the notion of behaviour to do this job, since it too is infected by content. If he would rather explain mere body-movements, then my argument does not succeed, but he must then give up all ambitions of the STM doing all that folk psychology can do and doing it better. That is, he must give up the moderate strategy.

In any case, perhaps Stich does not really favour this strategy. The schematic example he gives of an STM theory does, it is true, employ the idea of 'syntactic analogues' of beliefs and desires. But this is only meant to show that STM theories can do all the things that content-based theories can; Stich adds that 'nothing in the STM *requires* cognitive theories to mimic content-based theories' (p. 153; my emphasis). He also calls the view that the STM will follow the general contours of folk psychology 'implausible' (p. 182). In this at least he is right. So the STM should pursue the radical strategy.

The radical strategy says that an STM should be constructed without regard to the classifications of folk psychology. In order to answer the question, how does the STM establish which intrinsic properties are syntactic, the radical STM theorist says that we need not pay attention to the form of the belief/desire explanations of folk psychology, as the moderate strategy did. There is no need to let the future flourishing of cognitive science be hemmed in by the 'aboriginal language of content' and employ anything like the notions of a B- or D-state. STM theories will flourish unconstrained by such clumsy and imprecise notions.

In order for the radical strategy to succeed, it needs to distinguish

between two sorts of structure—causal structure in general, and syntactic structure in particular. It then needs to say why the mind has syntactic structure. A system—*e.g.* the solar system—may have a casual structure, in the sense that elements of the system obey certain laws that describe their causal interactions, without having syntactic structure. How can the radical STM distinguish between causal and syntactic structure, and then make a case for the mind's having the latter sort of structure?

Perhaps it can answer as follows. What it means for the mind to be a purely syntactic system is for it tc contain a finite number of primitive elements that combine to form wffs. What makes a formula a wff is that it is combined out of a number of the primitives according to the rules. The causal interactions of the wffs—which tokenings of wffs tend to cause which others—are determined by other rules that describe the (ideal) behaviour of the system. These two sorts of rules taken together constitute the STM for the system. We can formulate the second sorts of rules by looking at the relations between the inputs of the system (perceptions, non-intentionally characterised) the outputs of the system (body movements) and the relations between the wffs that we have established as we go along. Then we see which set of rules best accounts for these relations. What the syntactic properties of the wffs are is determined completely by the physical state of the brain and the nature of the syntactic rules of the theory. Thus, being P1 (an imagined property of the theory) is a purely syntactic property—its identity is determined by the rule that says that any wff that (say) is caused by P2 and causes P3 is, by 'theoretical definition' a P1. Naturally, to infer the rules from the complex actual behaviour of peoples' minds will require much careful study and idealisation—but, as Stich says, who ever thought that psychology was *easy*?

This response employs the strategy of theoretical definition: whether or not any property is syntactic will depend on which syntactic rules the theory employs. And whether or not the theory needs to appeal to this property rather than that depends on whether the explanatory and predictive power of the theory's generalisations is affected by choosing it. Before actually formulating an STM, this is all there needs to be said about what the syntactic properties of the theory are.

But this familiar theoretical strategy does not by itself make the properties in question syntactic. Consider this analogy.[14] A particular system of railway trains and tracks—say British Rail's 'Network South-East'—has a certain causal structure. What the trains do in this system is determined by two things—the layout of the tracks and the rules of the timetable. Trains can only run on the tracks, and when they run, and on which tracks, is fixed by the rules of the timetable. (This is an idealisation, of course, since all sorts of factors from outside the system can prevent the trains from

[14] Inspired by Schier 1986, p. 275, and Mellor 1989, pp. 51–7.

running according to these rules—human error, engine failure, weather conditions, industrial action.) Moreover, each train gets its identity in the system by its place in the system: what makes a particular train a token of the daily 16.05 from Cambridge to Liverpool St. is not anything (for example) about its appearance, but just the fact that it (ideally) leaves Cambridge at 16.05 daily and (ideally) arrives in Liverpool Street at 17.09.

Now imagine that we look at the behaviour of this system and we reason thus. 'Network South-East is a syntactic system! What this means is that it contains a finite number of primitives—carriages and engines—that combine to form *well-formed trains (wfts)*. What makes a train well-formed is just that it consists of an engine and at least one carriage, in any order—the engines can push or pull. The causal interactions of the wfts—stopping and starting at stations and signals—are determined by rules that describe the ideal behaviour of the system. These rules constitute the *timetable* for the system. We can infer these rules by looking at the relations between the inputs of the system (the people entering the stations) the outputs of the system (the people leaving the stations) and the relations between the wfts that we have established as we go along. What the syntactic properties of the wfts are is determined completely by the physical shapes of the track system, and nature of the syntactic rules in the timetable. Thus, being the 16.05 from Cambridge to Liverpool Street is a purely syntactic property—it is determined by the rule that says that any wft that leaves Cambridge at (approximately) 16.05 and arrives at Liverpool Street at 17.09 is, by "theoretical definition", the 16.05 from Cambridge to Liverpool Street. Naturally, to infer the rules from the somewhat rough and ready actual behaviour of the system will require much careful study and idealisation—but who ever thought train science would be *easy*?'

Why this analogy is absurd is not because there is something wrong with the idea of theoretical definition—there may be, but it is surely not anything peculiar to this analogy. What makes it absurd is the idea that there is anything syntactic about this causal system at all. To hold that it *is* syntactic involves a trivialisation of the idea of syntax, rendering it useless for distinguishing between the causal structure of the mind and that of Network South-East.

The fact that there is a level of description of the causal structure of a system, intermediate between the physical and the semantic or intentional, does not show that the system has a syntax. Of course, what the trains do is not determined by the laws of physics alone. But this does not make Network South-East syntactic, any more than the fact that what the economy does is not determined by the laws of physics alone makes the economy syntactic.

The question this section has posed is: why should we suppose that the mind has syntactic structure? The LT hypothesis has, as we saw, an answer to this question: intentional states have syntactic structure because of the nature of mental processes. But this is a thesis about the *semantics* of

thought. As Fodor has said, the question about whether there is an LT amounts to the question of whether

> mental states have a combinatorial [*i.e.* compositional] semantics: the kind of semantics in which there are (relatively) complex expressions whose content is determined, in some regular way, by the content of their (relatively) simple parts. (Fodor 1987, p. 138)

Thus for the LT hypothesis, the reason for assigning syntactic structure to internal representations is because these representations have a certain sort of *semantic* structure. But, *ex hypothesi*, the STM cannot appeal to this. Stich himself never addresses the question of what makes a property syntactic. So he does not offer a solution to the problem of distinguishing syntactic from causal structure generally. Of course, if mental states are states of the brain, then these brain states will have a neurological causal structure. But the point I am making is that there is no justification for the STM to talk of *syntactic* structure, once semantics are out of consideration. Without semantics, a theory of the mind is not entitled to distinguish syntactic structure from any other sort of causal structure: it is semantics that picks out syntax from the sea of non-semantics.

5. The Lesson of the Argument

So neither the moderate strategy nor the radical strategy yields an STM. The moderate strategy cannot succeed unless a semantically infected notion of behaviour is introduced. The STM cannot both mimic folk psychology and retain its independence from semantics. But on the radical strategy, there is no reason to think that the mind has syntactic structure at all.

There are still some loose ends. It might be objected (a) that I have tailored the notion of syntax to fit my argument; or (b) that although the argument undermines the STM, it does not support Fodor's Representational Theory of the Mind (RTM). I will answer these two objections in turn.

Perhaps Stich will say that the notion of syntax I have introduced into the discussion weighs the argument too heavily in my favour. In defending the idea of syntax as the causal mechanism for semantically compositional processes, I have imported semantics into the discussion in a way that disallows the very intelligibility of the STM at the outset. So perhaps he will wish to reformulate the STM using another notion of syntax that does not favour semantics as the ordinary notion does.

This does not look very promising. For one thing, the notion of syntax arises out of linguistics, the study of entities that are also semantically structured in an uncontroversial sense. So when applying the notion to the mind, the burden of proof is surely on those who think that the mind will not parallel language in the semantic respect too. For another, Stich

does not say what he means by syntax, so I am surely entitled to use the word in its standard sense, and then see how it may apply to the mind.

However, perhaps Stich can respond that the use of the word 'syntax' is immaterial. Perhaps he can employ a 'syntax-like' notion, the use of which 'can do justice to all of the generalisations capturable by quantifying over content sentences while avoiding the limitations that the folk language of content imposes' (1983, p. 157).

Of course, we should not make a fetish out of the word 'syntax'. But let us not forget what Stich's point was in introducing this notion. He explicitly rejects holistic views of intentional states in favour of mental sentence views (1983, pp. 23–8). The notion of syntax was supposed to provide a way of individuating mental states non-holistically. If Stich rejects my use of 'syntax', then he will have to find a way of picking out the structure of these states that is not semantic *and* not just physiological. And here the argument of Section 4 applies again, in a way that depends on no contentious assumptions about the meaning of 'syntax'. On the one hand, if STM theories are really going to do justice to all the generalisations of folk psychology, then they have to individuate the folk psychological explananda semantically. But on the other hand, if the syntactic properties of states of the brain are individuated solely by their neurophysiological relations, then the whole idea of an STM is vacuous, as the 'train science' analogy showed. Whichever word Stich uses, that is the dilemma, and the STM will be impaled on one of its horns.

Finally, I shall briefly consider the lesson of my argument to Fodor's RTM. It might be said that even if the argument of the last section is granted, the truth of Fodor's RTM does not follow. This is correct, if only because the falsity of the STM obviously does not entail that any mental sentence view is correct. The mind's internal structure may not be linguistic (see Stalnaker 1984, Chs. 1 and 2).

But I have only been arguing here for the conditional thesis: *if* there is an LT, then it must have semantics as well as syntax. And this is a claim with which Fodor will agree—it is, in a sense, another way of putting his slogan, 'No computation without representation'. However, if I am right, it seems Fodor is in a bind: for it is hard to see how he can maintain both that semantics is irrelevant to the theory of mind in the way methodological solipsism (as he formulated it in 1980[15]) requires it to be, *and* that the mind also might have a syntax in the straightforward sense I have described. For if there is a language of thought, and there cannot be a syntactic theory of the mind, then semantic notions are going to have to figure in the foreground of an account of the mind.

Stich himself appreciates this tension in Fodor's views (1983, p. 188). Fodor responds, correctly in my view, by distinguishing between general-

[15] 'Truth, reference and the rest of the semantic notions are not psychological categories' (Fodor 1980, p. 252).

isations over intentional states and the mechanisms which 'implement' instances of these generalisations:

> The vocabulary required to articulate the characteristic laws of a special science is—almost invariably—different from the vocabulary required to articulate the mechanisms by which those laws are sustained, the theory of the mechanism being pitched—to put it crudely—one level down. So the typical *laws* of psychology are intentional, and the typical *operations* of psychological mechanisms are computational. (1987, p. 166)

But this amounts to abandoning the idea that the semantic notions are not psychological. If the laws of psychology are intentional, then they will involve the semantic notions of intentional psychology: content, representation and truth.[16] This is a shift from Fodor's earlier position that semantic concepts have no place in psychology—a shift that is, in my opinion, to be welcomed.

So if the mind is a syntactic engine, it must be a semantic one too: there can be no syntactic theory of the mind. The choice for theories of intentional states, then, must be between those eliminative theories which eschew all talk of the mind, and those theories that are committed to semantic and intentional notions.

Centre for Philosophical Studies
Department of Philosophy
King's College
Strand
London WC2R 2LS

References

Bennett, J. 1976: *Linguistic Behaviour.* Cambridge: Cambridge University Press.
Blackburn, S. 1984: *Spreading the Word.* Oxford: Oxford University Press.
Blackburn, S. 1986: Finding Psychology. *Philosophical Quarterly*, 36, 111–122.
Block, N. 1981: *Readings in the Philosophy of Psychology Vol. II.* London: Methuen.
Chomsky, N. 1957: *Syntactic Structures.* The Hague: Mouton.
Chomsky, N. 1976: *Reflections on Language.* New York: Pantheon.
Chomsky, N. 1980: *Rules and Representations.* Oxford: Blackwell.
Churchland, P.M. 1981: Eliminative Materialism and Propositional Attitudes. *Journal of Philosophy*, 78, 67–89.

---

[16] A comprehensive critique of Stich would therefore have to undermine, by detailed examination of actual psychological theories, his confident claim that 'contemporary computational theories in psychology are STM theories' (Stich 1983, p. 153). I think this can be done, though it would be a task for another paper.

Churchland, P.S. 1986: *Neurophilosophy: Toward a Unified Science of Mind-Brain.* Cambridge, MA.: MIT Press.
Clark, A. 1989: *Microcognition: Philosophy, Cognitive Science and Parallel Distributed Processing.* Cambridge, MA.: MIT Press.
Clark, A. 1989a: Critical Study: J. Fodor, *Psychosemantics. Mind,* 98, 605–617.
Crane, T. 1989: *The Content and Causation of Thought.* Cambridge University, Ph.D. thesis.
Cummins, R. 1989: *Meaning and Mental Representation.* Cambridge, MA.: MIT Press.
Dennett, D. 1977: A Cure for the Common Code? *Mind,* 87, 265–280. Reprinted in Dennett 1978.
Dennett, D. 1978: *Brainstorms.* Hassocks: Harvester Press.
Dennett, D. 1978a: Three Kinds of Intentional Psychology. In R. Healey 1978, 37–61.
Dennett, D. 1982: Beyond belief. In A. Woodfield 1982, 1–95.
Dretske, F. 1981: *Knowledge and the Flow of Information.* Oxford: Blackwell.
Field, H. 1978: Mental Representation. *Erkenntnis,* 13, 9-61. Reprinted in Block (ed.) 1981, 78–114.
Fodor, J. 1975: *The Language of Thought.* New York: Crowell.
Fodor, J. 1980: Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology. *Behavioural and Brain Sciences,* 3, 63–110. Reprinted in Fodor 1981, 225–253.
Fodor, J. 1981: *Representations.* Hassocks, Harvester Press.
Fodor, J. 1983: *The Modularity of Mind.* Cambridge, MA.: MIT Press.
Fodor, J. 1985: Fodor's Guide to Mental Representation. *Mind,* 94, 76–100.
Fodor, J. 1987: *Psychosemantics.* Cambridge, MA.: MIT Press.
Fodor, J. 1987a: A Situated Grandmother? *Mind and Language,* 2, 64–81.
Grice, H.P. 1957: Meaning. *Philosophical Review,* 66, 377–88.
Higginbotham, J. 1988: Is Semantics Necessary? *Proceedings of the Aristotelian Society,* 88, 219–41.
Lewis, D. 1983: New Work for a Theory of Universals. *Australasian Journal of Philosophy,* 61, 343–77.
McGinn, C. 1982: The Structure of Content. In A. Woodfield 1982, 207–258.
Mellor, D.H. 1981: *Real Time.* Cambridge: Cambridge University Press.
Mellor, D.H. 1989: How Much of the Mind is a Computer? In P. Slezak and W.R. Albury 1989, 47–69.
Peacocke, C. 1983: *Sense and Content.* Oxford: Oxford University Press.
Pylyshyn, Z. 1984: *Computation and Cognition.* Cambridge, MA.: MIT Press.
Quine, W.V.O. 1960: *Word and Object.* Cambridge, MA.: MIT Press.
Schier, F. 1986: The Withering Away of the Cognitive State. *Philosophical Quarterly,* 36, 260–78.
Schiffer, S. 1987: *Remnants of Meaning.* Cambridge, MA.: MIT Press.
Sells, P. 1985: *Lectures on Contemporary Syntactic Theories.* Stanford: CSLI.
Shoemaker, S. 1980: *Identity, Cause and Mind.* Cambridge: Cambridge University Press.
Slezak, P. and Albury, W.R. 1989: *Computers, Brains and Minds: Essays in Cognitive Science.* Dordrecht: Kluwer.
Stalnaker, R. 1984: *Inquiry.* Cambridge, MA.: MIT Press.
Stampe, D.W. 1977: Toward a Causal Theory of Linguistic Representation. *Midwest Studies in Philosophy Volume 23.* Minneapolis, MN.: University of Minnesota Press, 42–63.

Stich, S.P. 1983: *From Folk Psychology to Cognitive Science*. Cambridge, MA.: MIT Press.
Tarski, A. 1944: The Semantic Conception of Truth. *Philosophy and Phenomenological Research*, 4, 341–75.
Wittgenstein, L. 1953: *Philosophical Investigations*. Oxford: Blackwell.
Woodfield, A. 1982: *Thought and Object*. Oxford: Oxford University Press.