

A shorter version will appear in
Boston Studies in the History and Philosophy of Science
(Dordrecht: Kluwer)

[This is a working draft of a chapter of a book manuscript long overdue,
somewhat incomplete and unpolished: please read it accordingly.]
Comments/criticisms/corrections are most welcome.

**SYNTAX, CONTENT, AND FUNCTIONALISM:
WHAT IS WRONG WITH
THE SYNTACTIC THEORY OF MIND**

MURAT AYDEDE

The University of Chicago
Department of Philosophy
1010 East 59th Street
Chicago, IL 60637

(E-MAIL: m-aydede@uchicago.edu)

1993/98

ABSTRACT. I argue that Stich's Syntactic Theory of Mind (STM) and a naturalistic narrow content functionalism run on a Language of Thought story have the same exact structure. I elaborate on the argument that narrow content functionalism is either irremediably holistic in a rather destructive sense, or else doesn't have the resources for individuating contents interpersonally. So I show that, contrary to his own advertisement, Stich's STM has exactly the same problems (like holism, vagueness, observer-relativity, etc.) that he claims plague content-based psychologies. So STM can't be any better than the Representational Theory of Mind (RTM) in its prospects for forming the foundations of a scientifically respectable psychology, whether or not RTM has the problems that Stich claims it does.

TABLE OF CONTENTS

1. Introduction

2. The Problem Space according to Stich

3. What Is STM?

4. Some Curious Aspects of STM

5. The Alleged Superiority of STM

5.1 How STM Is Supposed to Be Superior

5.2 The Parallel Disadvantages of STM-Style Theories

5.2.1 The Referential Dimension and Narrow Content

5.2.2 Doxastic Similarity Dimension and the Holism Problem

5.2.2.1 Stich's "Holism"

5.2.2.2 Holism and STM

5.2.3 Dimension of Narrow Content Pattern Similarity

6. The NCA and the Type Individuation of Brain Sentences

6.1 Troubles with Content Functionalism

6.2 Connections to Stimuli and Behavior

7. Why a Purely Syntactic Psychology Cannot Get off the Ground

8. If Cognition Is Computational, How Can Psychological Laws be Intentional?

REFERENCES

1. Introduction

There is a thesis often aired by some philosophers of psychology that syntax is all we need and there is no need to advert to intentional/semantic properties of symbols for purposes of psychological explanation. Indeed, the worry has been present since the first explicit articulation of the so-called Computational Theory of Mind (CTM). Even Fodor, who is the most ardent defender of the Language of Thought Hypothesis (which requires the CTM), has raised worries about its apparent consequences. The worry can be put in the form of a question, which Fodor called the “Eponymous Question” alluding to the title of a chapter in his (1994) book: If cognition is computational, how can psychological laws be intentional? This question has been haunting people working in the field at least since the publication of a paper in 1978 by Stich in which he gave his celebrated “autonomy argument”. Then, as everybody knows, came Fodor’s notorious “Methodological Solipsism” in 1980, in which he argued for the *formality condition*: namely, thought processes are causal sequences of symbol tokenings in one’s language of thought (LOT) and the causal processes are sensitive only to the syntactic/formal properties of its symbols. Hence, he argued against what he has called a “naturalistic psychology,” i.e. a psychology whose laws essentially advert to broad semantic properties of mental states they cover. The alternative, rationalist psychology, according to Fodor, was to advert only to formal characteristics of symbols, of which Fodor conceived as narrow computational roles of LOT symbols.

Stich’s 1983 book, *From Folk Psychology to Cognitive Science*, was the culmination of the worries presented as a sustained argument against the possibility of a scientific *intentional* psychology (along with the common sense belief-desire psychology), and for a syntactic way of doing psychology, i.e. for his much discussed Syntactic Theory of Mind (STM). He defended an eliminativist stance: STM involves the elimination of all intentional idioms proposed to be used in a scientific enterprise, hence envisions a scientific psychology free of semantics. STM has been around for about two decades now. It has generated a lot discussion because it has usually been taken to articulate the paradox alleged to underlie the Language of Thought Hypothesis (LOTH), which was to vindicate intentional folk psychology through computationalism. For this reason, I will concentrate on Stich’s book in what follows, and argue that the worries are altogether baseless, that a computational theory needs a semantic individuating scheme to get off the ground, and that the envisioned alternative, i.e. a pure CTM or STM is a non-starter and cannot do the required job. Although there are probably few actual adherents of STM nowadays or the vision it has of a scientific psychology, to my mind, no one so far has been able to conclusively refute the theory. Indeed, as recently as 1994, Fodor raised the worry and tried to answer it by showing the feasibility of its alternative, and not by directly attacking the syntacti-

calist claim. In what follows, I intend to refute once and for all not only the STM but also the kind of semantic-free psychology it envisions, thus answering Fodor's Eponymous Question. Along the way, I will also attack narrow content functionalism as a naturalistic semantic program, since, as we'll see, it will turn out to have a very close relation to STM.

In his book, Stich has argued basically for two claims. First, the application conditions of such intentional common sense predicates as 'believes that *P*' and 'desires that *Q*' are essentially vague, context-sensitive, observer-relative, and thus are not suitable to be used as stable projectible predicates in the vocabulary of a scientific psychology. In particular, since, according to Stich, observer relativity partly stems from the fact that content ascriptions are essentially based on similarity judgments along different dimensions (see below) between the ascriber and the ascribee, a consequence of observer relativity of ordinary content ascriptions is that a certain form of parochialism will profusely infect our psychological theories if we insist on having a content-based psychology, which, according to Stich, essentially relies on ascribing such contents to agents covered by its generalizations. This means that content-based psychologies are bound to miss many important generalizations about the psychology of children, exotics, perceptually or cognitively handicapped people, higher animals, etc., since any content ascribed to these will necessarily reflect how cognitively similar the ascribee is to the ascriber.¹ In short, Stich thinks that content-based psychologies won't make respectable science.

His second general claim is that we don't *need* to advert to the content of mental states in doing psychology, "syntax" will be enough, and we had better advert only to the syntactic properties of mental states if we don't want to miss any psychological generalizations: STM offers a paradigm that has all the virtues of content-based psychologies and none of its vices.

These two claims are relatively independent of each other: in particular, the truth of the latter does not depend on the truth of the former. If Stich is right in his second claim, then the falsity of his former claim, i.e. his characterization of contentful mental states as scientifically problematic posits, would imply that content vocabulary is at best otiose in doing scientific psychology. It is therefore important to see whether Stich is right in his second claim. So my attack will be on his second claim, and I will leave the discussion of the first aside in this paper. In what follows I will argue for three basic claims.

The first is that when we see more clearly the essence of what STM is supposed to be, as presented by Stich, the claimed superiority of STM over content based psychologies totally disappears. Put differently, I will be arguing for a conditional claim: *if* Stich is right in his claim

¹ Stich does not distinguish between beliefs and belief ascriptions, but takes the consequences of his analysis of the latter equally applicable to the former. As also indicated by Fodor (1987) and Devitt (1996), I think this is a major mistake. But since my concern will be what Stich thinks ordinary belief individuation involves, I will not dwell on this unfortunate slip

that content-based psychologies have the disadvantages he enumerates, *then* STM-style theories have exactly the parallel problems; so it is false that the STM framework is scientifically superior to content-based psychologies as conceived by Stich. Therefore, Stich will lose his primary motivation to promote STM. This is the topic of § 5. Although I think Stich is wrong about his first claim, I will say very little about it in what follows.

Secondly and more boldly, I will argue in § 6 that STM can't do the required job: it lacks the necessary resources to type individuate *particular* psychological states *qua* mapped onto *particular* "syntactic objects" as Stich puts it.

In connection to this last point, I'd like to bring out a curious and *prima facie* puzzling aspect of Stich's second claim, i.e., the claim that the only scientifically viable paradigm in psychology is and ought to be STM, by relating it to some worries that Fodor has about the nature of semantics. Fodor, as is well known, has opted for a purely denotational theory of meaning. He has various reasons for this, but the most important one is that the alternative, the functional role semantics or content functionalism in general, seems to have consequences that are destructive of scientific intentional psychology. The problem is supposed to be that any bit of functionalism in semantics inevitably leads to holism, or else has theoretical commitments that are very dubious on independent grounds.

Now, Stich, on the other hand, does think that *one of the reasons* why intentionally characterized mental states like beliefs and desires are ill-suited for scientific purposes is that their individuation conditions are inherently holistic, and therefore, intentionally characterized states are a mess beyond repair to serve as stable law-instantiating theoretical posits. So Stich offers his STM within which scientific psychology ought to be developed. But his STM, as I will present in a moment, is committed to what he calls the Narrow Causal Account (NCA) of typing brain states, which makes it a purely functionalist theory.

So, in a nutshell, here is the puzzle: Whereas Fodor thinks that it is because functionalism cannot individuate contents we have to develop denotational approaches in semantics if we want to have a scientific intentional psychology, Stich, on the other hand, thinks that it is because content is inherently holistic (among its other defects) that we have to develop purely functionalist but semantic-free psychological theories if we want psychology to be scientifically respectable, hence his STM!

Intuitively, we may ask: If Fodor is generally right about what he thinks of functionalism in semantics, how can Stich think that his purely functionalist STM can be free of similar or parallel defects? Or, conversely, if Stich is right about what he thinks of functionalism in its

of Stich for what follows.

capabilities to type brain states in a scientifically acceptable way, why can Fodor not utilize functionalism in his story about fixing the semantics of mental states?

Putting the question in this way might seem misleading to many. For one might object: Fodor's worry is about the role and place of functionalism in fixing the semantic content of Mentalese expressions, but Stich couldn't care less about it because his brand of functionalism has nothing to do with semantics; rather, Stich's claim is that the syntax of "sentences" realized in the brain should be fixed purely functionally and syntax is all we need for a scientific psychology.

My second claim will take up exactly this issue. I'll argue that given a certain picture of the cognitive mind in broad outline, a certain brand of functionalism, namely the one that is committed to the NCA, is equally problematic both in fixing the content of psychological states and in fixing their "syntax," especially as understood by Stich and, in certain of his moods, by Fodor as well. In arguing for my claim, my strategy will be as follows. I will show that STM is a purely functionalist theory whose structure is exactly parallel to content functionalism of the NCA variety (§ 3). Then I will show that such a content functionalism cannot type-individuate the particular semantic contents that are to be assigned to central psychological states (like beliefs and desires) on the basis of their narrow causal role (§ 6). I will then argue that the burden of the proof is on Stich to show how STM is supposed to accomplish the exactly parallel feat of assigning particular syntactic objects to psychologically relevant brain states on the basis of their narrow causal profile. I will conclude that STM can't type-individuate the particular brain states *qua* mapped to particular syntactic objects. So, STM, contrary to Stich's advertisement, will turn out to be incapable of providing a solid framework within which a scientifically respectable psychology can be pursued.

Thirdly, I will argue in § 7 that the STM-theorist is, at any rate, committed to intentional vocabulary at some stage of theory construction. Put differently, if the STM strategy is taken to claim, as Stich seems to intend, that it is possible and advisable to develop psychological theories without using any intentional scheme whatsoever no matter what the stage of theory construction is, then STM is false: psychological theory construction cannot get off the ground if the strictures of STM are firmly complied. Admittedly, this is an issue that belongs to the "context of discovery", but I believe it is still instructive to see why STM has to rely on intentional vocabulary at least in the initial stages of theory construction. For, this will reveal how wrong-headed Stich is both in his conception of syntax and the place of functionalism in psychology.

I will end (§ 8) by moralizing on Stich's failure, and point out that if STM (= Narrow Causal Account of typing symbols over which computation is defined, as I will show) and type-type identity theory are false, as I will argue, then a content-based psychology (= intentional

psychology) is practically mandatory. Hence, if cognition is computational, psychological laws have got to be intentional!

Since all my arguments crucially depend on what exactly STM is, I will present it in a way that its purely functionalist structure becomes explicit. This is the job of § 3. In particular, I will use a procedure similar to the one developed by Brian Loar (1982a) in his presentation of his own content functionalism. Indeed this is how I want to establish the exact structural parallelism between STM and a narrow content functionalism.

However, before embarking on my criticism, I need to say a little about how Stich views the problem space within which he criticizes content-based psychologies and thus motivates his own alternative, STM. In particular, it will turn out that the exact way in which Stich motivates his STM is very important, since my arguments against STM partly rely on his own strategy. So we will need this general background. This is the job of the next section, § 2.

In § 4, I will comment on some curious aspects of STM, especially on what apparently makes STM a “syntactic” theory. I will argue that in one reading, STM has nothing to do with “syntax”. In particular, I will present the reasons for why the notion of syntax Stich has in mind is not the notion of syntax that the Language of Thought Hypothesis (LOTH) requires. Hence, we will see that, despite the widely held opinion to the contrary, STM has very little to do with the Computational Theory of Mind (CTM).

2. The Problem Space according to Stich

Stich takes what he calls the Mental Sentence Theories as his starting point, and assumes their basic framework throughout his discussion. After a lengthy presentation of a Fodor-style LOTH he raises the following problem:

for a Fodor-style account of belief sentences to hang together, we must have some workable notion of what it is for two distinct people, speaking different languages, to have in their heads distinct tokens of the same sentence type. (Stich, 1983: 43-4)

On behalf of Fodor, he offers three possible solutions: One of them is the *Narrow Causal Account* (NCA), according to which two sentence tokens count as type identical if and only if (iff) they have the same narrow causal/functional role. Since this is going to be of some importance, let me elaborate on it a bit. According to Stich, “[t]o adopt this view of... psychology is to exclude any reference to noncausal relations... There can be no mention of a subject’s social setting, natural environment, or personal history, nor of the psychological characteristics of other people” (1983, p. 22). This is what makes this kind of individuation *narrow causal*. It is narrow be-

cause the causal role in question is defined in terms of generalizations that detail nomological connections among *proximal* stimuli, behavior (like motor commands) and other central cognitive states. Moreover, the causal relations are given by a set of *counterfactual supporting* generalizations. Thus, for a mental state of an individual to count, say, as the belief that P, it is not necessary that the state *actually* play a causal role in the individual's mental economy; all that is required to be true of the state is that it *would play* a certain causal role if some other conditions specified in the generalizations *were to obtain*. So the notion of functional/causal role of a mental state should be so understood as to include the *potential* causal interactions that the state would enter. Accordingly, two mental states of two distinct organisms count as of the same type if their potential causal interactions are the same, namely, if they are covered by more or less the same generalizations, despite the fact they may differ quite radically in their actual etiologies. Finally, the generalizations in question are hedged by *ceteris paribus* clauses.

The second account is what Stich calls the *Semantic Account* (SA) according to which two sentence tokens count as type identical iff they have the same semantic content. The third one is what might be called the *Quasi-Physical Account* (QPA) according to which the sentence tokens are of the same sentence type iff their quasi-physical properties, their shape, so to speak, are the same.² After quickly dismissing the QPA as hopeless, he makes the following remark:

[The] interesting question is how causal accounts and content accounts compare with each other. Do they categorize mental tokens differently, or do they inevitably come out with the same categorization? On this issue, opinions divide. According to Fodor the two sorts of classification schemes coincide, "plus or minus a bit." Indeed Fodor sees this as "the basic idea of modern cognitive science." Any thoroughgoing [i.e., content] functionalist in the philosophy of mind will also end up on this side of the divide. On the other side, denying that causal and content accounts converge, are Field, Lycan, Perry, McDowell, and the truth. (1983: 48–9)

Here Stich conceives of the Semantic Account as fixing the type identity of mental sentences according to their broad content. Twin-earth cases show that functionally identical twins may differ in the broad content of their mental states. However, this is not going to be very impor-

² The reason I call this Quasi-Physical Account is that formal/syntactic properties in this sense can still be abstract, higher-order physical properties. For instance, a complete specification of the shape of a symbol would be the specification of its formal/syntactic properties in this sense. But shapes are still abstract entities. Shapes of letters, for instance, can be realized in a variety of physical media. Just think of the letter 'A' inscribed in sand, wax, etc. If that is right, syntactic properties of symbols in this sense can be multiply realized without being functionally defined. This is why I call them quasi-intrinsic.

tant for what follows. Since, for many people in the field, narrow content is a construct out of broad content, Stich has the same line of argument against narrow content.³

Here is Stich's argumentative strategy. Stich thinks that if folk psychology is to be scientifically vindicated through some version of a mental sentence theory, the Semantic Account of typing mental sentence tokens is indispensable. He then proceeds to show that the NCA and SA come up with radically different taxonomies. The way he does this is idiosyncratic. He constructs a series of thought experiments that are supposed to intuitively show that folk judgments about how to classify certain mental states do radically differ from the way the NCA would type them. Then relying on what these thought experiments seem to show, he proceeds to give an account, or rather a "descriptive analysis", of folk conception of belief as a paradigm case of a contentful mental state, i.e. as a paradigm case of a mental state typed according to SA.

According to Stich's analysis, the "content identity" of beliefs that is thought to be assumed by folk psychology is a myth. On the basis of the evidence he claims to have collected through his thought experiments, he claims that the notion of content according to folk psychology is such that it is only a similarity measure along three different dimensions that the folk implicitly assume. One dimension of similarity between contents is the *functional or causal-pattern* of contentful mental states: "A pair of belief states count as similar along this dimension if they have similar patterns of potential causal interaction with (actual or possible) stimuli, with other (actual or possible) mental states, and with (actual or possible) behavior" (Stich 1983, pp. 88–9). The second dimension draws on the *ideological (doxastic) background* of the agents. Since these can greatly vary from person to person, the relation between two beliefs in two different people can only be a matter of similarity: "The ideological similarity of a pair of beliefs is a measure of the extent to which the beliefs are embedded in similar networks of belief" (p. 89). The third dimension of similarity measure concerns the *reference or truth-conditions* of beliefs. Since they are dependent on the speakers' linguistic community, social imbeddings, the causal history of the use of terms, as well as the physical and cultural environments of the speakers/believers, etc., the reference will vary as these vary without necessarily affecting the functional role of a mental state. To the extent that these factors are similar, to that extent the contents of beliefs will be similar. Stich thinks that this is essentially what the SA of typing mental sentences comes down to. Stich, of course, needs such an analysis because he has to know what it is exactly that a mature cognitive science threatens to eliminate and why.

³ In his (1991) Stich argues against Fodor that narrow content taxonomies will differ from the narrow causal taxonomies, which he calls taxonomies according to the fat syntax of mental sentences. The problem, according to Stich, stems directly from the Semantic account itself, narrow or wide. See § 5.2 below.

It is now relatively easy to see how the two taxonomic schemes diverge. The NCA can capture only the *causal pattern similarity* dimension assumed in the SA. It can't be sensitive to the other dimensions. Stich concludes that "the mental sentence theory of belief, if fleshed out with a narrow causal account of belief, just does not comport with our workaday folk psychological notion of belief — it is not an account of belief, as the term is ordinarily used" (1983, p. 49).

If the two taxonomic schemes differ, what scheme should a scientific psychology adopt? Stich argues that adopting the SA is ill-advised, because mental states so typed will make bad science since a semantic taxonomy would only provide the psychologist with a theoretical vocabulary whose application is vague, unstable, context-sensitive, and observer relative. (I will return to this issue in § 5.1) Who would want such a science, Stich argues, especially if there is a clear alternative that is free of such defects? According to Stich, the alternative is a psychology whose taxonomic scheme is based on the NCA. This is the STM paradigm. Hence Stich's main conclusion: if a mature cognitive science is and ought to be committed to the NCA (STM), then folk psychological intentional notions like beliefs and desires are likely to be eliminated.

This is how Stich motivates and argues for his STM. It is therefore very important to see whether Stich is right in his claim that the STM paradigm is really superior in any of the respects in which he criticizes content based psychologies. As I advertised, I will argue that Stich is wrong (§ 5).

There are a number of curious features in Stich's discussion. In particular, Stich does not distinguish between having a revisionist intentional psychology (hence, one committed to some version of SA) and scientifically vindicating folk psychology in its entirety. This is especially apparent in his identification of SA with the folk conception of contentful states. I will take up this issue in § 5.2.

Stich's indictment of intentional psychology is essentially based on his "descriptive analysis" of the folk notion of a belief. His discussion, however, seems to contain some curious confusions. In particular, as I said, Stich does not distinguish between beliefs and belief ascription. From a very idiosyncratic account of the latter he draws an account of the former. I don't think that there is any good reason to buy his account, but I won't argue against it here.

Let's now see what sort of approach STM is.

3. What Is STM?

According to Stich, the core idea of STM can be captured in the following way:

the cognitive states whose interaction is (in part) responsible for behavior can be systematically mapped to abstract syntactic objects in such a way that causal interactions among cognitive states, as well as causal links with [proximal] stimuli and behavioral events, can be described in terms of the syntactic properties and relations of the abstract objects to which the cognitive states are mapped. More briefly, the idea is that the causal relations among cognitive states mirror formal relations among syntactic objects. (1983: 149)

Stich here considers two networks, one of which is the network consisting of the causal relations among brain state types, proximal stimuli and behavioral events. This network is supposed to be mirrored by another network expressed by a syntactic psychological theory T. This theory consists of at least three kinds of generalizations: (1) the ones that nomologically connect proximal stimuli to B-states (belief-like states) with *particular* syntactic objects mapped to them, (2) the ones that describe causal relations among B-states and D-states (desire-like states), and (3) the ones that nomologically connects B- and D-states to motor-gestures. Following Michael Devitt (1990), I will call these kinds of generalizations: I-T, T-T, and T-O generalizations respectively.⁴

If we want to put T into some canonical form, we may write out T as a single conjunctive sentence, replacing all the occurrences of the theoretical predicates such as “x has a B-state mapped to ϕ_1 ” and “x has a D-state mapped to ϕ_1 ” with expressions of the form:

x is in (some member of) B(ϕ_1),

x is in (some member of) D(ϕ_1).

B and D are functions (in the set theoretic sense) that map a particular syntactic object, which the theorist had already specified for the job at hand, onto the set of x’s first order physical state types that have the functional role that T associates with that syntactic object. We may now express T in the following way:

(i) T[s₁, s₂,..., B(ϕ_1), B(ϕ_2),..., D(ϕ_1), D(ϕ_2),..., b₁, b₂,...]]

⁴ Generalizations detailing the causal relations between proximal input events and T-states (thought-like states), among T-states, and between T-states and proximal output events. See Devitt (1990).

where s_i 's are proximal stimulus types and b_i 's behavioral event types (motor gestures), and σ_i 's are specific syntactic objects.

Roughly, this is the form an STM theory would take. Let us now see how STM is committed to the NCA of typing brain states hypothesized by the theorist, i.e., how we can get their explicit functional definitions.

From (i) it is easy to get the Ramsey sentence of T by quantifying over the functions B and D :

$$(ii) \quad (\exists f_1)(\exists f_2) T[s_1, s_2, \dots, f_1(\sigma_1), f_1(\sigma_2), \dots, f_2(\sigma_1), f_2(\sigma_2), \dots, b_1, b_2, \dots]$$

We can now get the explicit functional definition of B :

$B = \text{df.}$ The function f_1 , such that there is a function f_2 , such that the two uniquely satisfy ' $T[s_1, s_2, \dots, x_1(\sigma_1), x_1(\sigma_2), \dots, x_2(\sigma_1), x_2(\sigma_2), \dots, b_1, b_2, \dots]$ '.

Similarly for the definition of D .

Although this is the formal procedure to get the explicit functional definitions of B and D , what we really want is explicit functional definitions of ' $B(\sigma_i)$ ' and ' $D(\sigma_i)$ ' for each i . The intuitive idea is this. Notice that in this formalism the existential quantification in getting the Ramsey sentence is over certain functions that map distinct syntactic objects to distinct sets of an organism's first order physical states. Here, in a certain sense, syntactic objects are exploited as external indices that pick up certain states of an organism that have distinct functional roles as specified by theory T . Each specific syntactic object in virtue of its distinctive place in T 's generalizations specifies a unique functional role that the two functions B and D then map onto the underlying physical states of the organism. The syntactic objects may be viewed to be indices that are external to the underlying states (but see below). They only function to pick up certain states with certain functional roles. Intuitively, we may extract the functional definition of $B(\sigma_i)$ for each i in the following way: since σ_i in the domain of B , in virtue of its place in T , is supposed to pick out a unique functional role that may be indexed by F_i ,

$B(\sigma_i) = \text{df.}$ the set of first order states that have F_i as determined by T .

Similarly for D and for each particular σ_i .

Now Stich does not present his STM in this way. Here I have used a procedure very similar to the one developed by Brian Loar (1982a) in his presentation of his own content functionalism. This is not accidental of course. In fact, this is the point. For, as should be obvious, Stich's STM, *structurally* at least, is nothing but a de-intentionalized version of Loar's content functionalism, except that Loar takes the causal role of "observational" beliefs to be fixed on the basis of *distal* stimuli. Where Stich uses abstract syntactic objects, Loar uses ("fine-grained") propositions, intentional objects *par excellence*. The type identity of specific abstract syntactic objects is given by their place in the theory. This is the way they are purely functionally defined according to their narrow causal profile.

In fact, the similarities between STM and Loar's content functionalism are, in one respect, stronger than that. Loar uses propositions in the initial stage of getting the functional theory first. (And, for good reasons, see § 7.) He then proposes a procedure by which all the propositions are replaced by purely formal expressions. The theory in this ultimate form is structurally almost an STM! Loar, of course, is no eliminativist. His aim is to naturalize intentionality by offering a sophisticated functionalist theory. So he thinks at some stage he should get rid of the intentional objects like propositions he had initially used. Once the theory is completed, it is supposed to provide sufficient (and, necessary?) conditions for a mental state to have a semantic content, which can ultimately be specified without using any intentional terminology. This is his strategy, and as far as it goes it is perfectly kosher. But if I am right in what I am going to say, it does not go very far at least in its narrow version (see § 6).

4. Some Curious Aspects of STM

My presentation of Stich's STM may be taken to be tendentious. I presented it as a purely functionalist theory and said that the abstract syntactic objects, which the brain states are mapped onto, may be viewed as indices that are totally external to the underlying first order brain states. But, STM is supposed to be a formal/syntactic theory very similar to the Computational Theory of Mind (CTM) Fodor has developed and defended. STM is supposed to be a de-intentionalized version of what Stich calls Mental Sentence Theories. Indeed STM has been taken in this way in the literature by its friends and foes. But if my presentation is right, STM is not in fact theoretically committed to there being syntactically complex "sentences" literally realized in the brain. If so, how could it be very similar to Fodor's CTM?

In fact, there is no mystery here. STM as presented by Stich is indeed not committed to there being sentences realized in the brain, as Stich himself acknowledges:⁵

⁵ See also pp. 78–9, (1983), where Stich writes: "mental sentence theorists typically leave

It is not, strictly speaking, required for an STM theorist to view hypothesized neurological state tokens as mental sentence tokens, though talking of them in this way is often an all but unavoidable shorthand. (1983:152)

This is curious but actually quite understandable. Remember Stich's question about how the tokens in different heads can be individuated as of the same sentence type. His solution is the NCA. But the NCA requires a theory first in which syntactic expressions figure as theoretical terms in the generalizations. However, once we have such a theory, it is easy to define the syntactic expressions functionally à la Loar. But once we do that, the question whether the referents of such expressions do really have syntactic structure somehow realized in the brain becomes secondary and at best an open empirical question. For, if the functional theory is true, we can do everything we want that the Mental Sentence version of the theory can do.⁶

So STM as a purely functionalist theory is not committed to a semantic-free LOT. On the other hand, of course, whatever CTM is, it cannot be neutral with respect to the question of whether there are syntactically complex sentences realized in the brain. CTM should be so formulated that it essentially entails a positive answer to this question. The problem in fact stems from the widely shared conviction that the type identity of brain sentences can and should be given in terms of the NCA (for some, as well as in terms of other ways like the SA).⁷ In § 6, I will argue that this can't be done. So there is at least this dissimilarity between STM and CTM: whereas STM is non-committal about there being brain sentences, CTM, whatever it is, is essentially committed to it.

Having made the point, however, I want to talk of STM most of the time as if it's concerned with the functional individuation of syntactically complex brain sentences. Not only because, as Stich says, this is an all but unavoidable shorthand, but also because I want to see whether Mentalese expressions can be individuated on the basis of the NCA if the LOTH is

the notion of an internalized sentence token as little more than a metaphor. And it may well turn out that when the metaphor has been unpacked, it claims no more than that beliefs are relations to complex internal states whose components can occur as parts of other beliefs." Here, it is not clear what the contrast Stich is trying to convey is supposed to be.

⁶ Indeed, this was the very point of Brian Loar in his polemical article (1982b) written against Fodor's LOTH. He says that from a philosophical point of view his non-committal content functionalism is weaker than the LOT version of it and thereby should be preferred. He does not reject the LOTH, but he claims that its motivation cannot be due to its having more explanatory and predictive power. For, with respect to these, his pure functionalism is equally good. Loar views LOTH as a scientific hypothesis, and as such, he leaves it as an open question.

⁷ Fodor used to think that way too.

true. So, in what follows, I will assume the framework of Mental Sentence Theories, and often treat STM in this form.

So far we have been talking about the functionalist nature of STM, and thus its commitment to the NCA of typing brain states. But what does this have to do with syntax? More particularly, how does Stich conceive of syntax when he talks about the syntactic type identity of brain sentences? Or, what makes his theory a “syntactic” theory? To this last question he answers in the following way:⁸

We would have no reason to view brain states as syntactically structured unless that structure can be exploited in capturing generalizations about the workings of mind/brain’s mechanisms. Attributing syntactic structure to brain state tokens — assigning them to syntactic types — is justified only if some interesting set of causal interactions among those tokens is isomorphic to formal relations among abstract syntactic objects. (1991: 244)

Notice that if Stich is right about this, Fodor can’t have any reason for postulating a separate computational level in which intentional laws of psychology are implemented.⁹ In particular, what is puzzling about Stich’s answer is that he doesn’t mention at all the Turing legacy which is the main driving force behind Fodor’s insistence that the computational story, according to which thought processes are defined over the formal/syntactic properties of representations, is our only plausible story about how semantically coherent processes can be physically/mechanically possible. Stich’s interest seems not to be in computationalism. This is understandable to a certain extent. For Stich doesn’t think that there are any semantically coherent thought processes that need the attention of science because he doesn’t think that there are any states with semantic content. Put this aside. He has a different line of answer.

When he talks about the syntactic type identity of brain sentences, he has a “rich” notion of syntax, according to which mere difference in lexical items (e.g. “Tully was bald” versus “Cicero was bald”, or “*Fa*” versus “*Fb*”) is enough to make the sentence tokens belong to differ-

⁸ He makes the same point in his (1983): “The core idea of the STM — the idea that makes it syntactic — is that generalizations detailing causal relations among the hypothesized neurological states are to be specified indirectly via the formal relations among the syntactic objects to which the neurological state types are mapped” (p.151).

⁹ There may be many separate implementational levels of course. But the point is that Stich does not mention any of these, and treats the nomological relations among mental states all at the same level, i.e., syntactically. If that were right, it would mean that psychological laws need not advert to intentional properties of representations, which is exactly the issue here.

ent syntactic types.¹⁰ In particular, for Stich, the criterion according to which two sentence tokens in two different heads count as of the same type is a syntactic criterion. But since this criterion is captured by the NCA, the syntactic type identity of brain sentences is a matter of functional identity:

when mental states are viewed as tokens of syntactic types, the functional profile exhibited by a mental state can be equated what we have been calling its formal or syntactic properties. (Stich, 1983: 190)

So it seems that, according to Stich, the very postulation of complex semantic-free sentences realized in the brain whose “syntactic” type identities are given purely functionally is what makes Stich’s theory a syntactic theory. I would like to point out that such an understanding of the notion of syntax is very different from the one that is needed for a Fodorian Computational Theory of Mind: what is required for the LOTH is a combinatorial syntax that fixes the logical form of expressions.¹¹ The important question I will address in § 6, however, is whether the type identity of brain sentences can be given in terms of their narrow causal profile, whatever it is called.

Now that we have set the stage, let us see whether the STM paradigm is any superior to content-based psychologies. In the following section I will present my argument for my conditional claim.

5. The Alleged Superiority of STM

From Stich’s analysis of folk conception of belief individuation it follows that predicates like ‘believes that P’ (1) are vague and unstable, (2) depend on a (observer-relative) similarity matrix along three different dimensions for their applicability, and (3) their application involves

¹⁰ Note that Stich’s claim is stronger than merely saying that lexically different sentence tokens have the syntactic property of *being different*. His claim is that they belong to different *specific* syntactic categories.

¹¹ Stich, in his (1991) article, calls the type identity of sentences that gets fixed on the basis of their narrow causal profile their “fat syntactic” identity. This is contrasted to their “skinny syntax”. The latter is supposed to be fixed only by the T-T generalizations: no causal relations to proximal stimuli and behavioral events can be used in the individuation of sentences. Stich insists that it is the fat syntactic type identity that would do the work for STM-style theories. As I said, I will argue that the NCA cannot fix the type identity of mental sentence tokens whether or not what gets fixed is their (fat) “syntactic” type. Devitt (1990) has argued that even if their type identity can be so fixed, what gets fixed would be their narrow semantics not their syntax. Devitt’s discussion also contains a very helpful criticism of Stich’s notion of syntax.

many unnecessary “fine-grained distinctions which contribute nothing in the explanation [and prediction] of behavior.” From (2), it also follows that there are likely to be many important cognitive generalizations that will not be stateable in terms of such predicates. So a content-based psychology will inherit all of these limitations. In contrast, the STM style theories, Stich claims, will have none of these.

In this section, I will argue that *if* Stich’s criticism of content-based psychologies is right then exactly parallel problems equally plague STM. But for this, we first need to see, exactly, how Stich argues for the superiority of STM. In other words, since my claim is conditional, we need to see in some detail what makes its antecedent true according to Stich and why he thinks that STM is free of similar problems.

5.1 How STM Is Supposed to Be Superior

In discussing how STM theories will succeed where the content theories fail, Stich again uses the thought experiments he has considered in showing how the content taxonomies radically differ from the ones based on the NCA. Much of the difference stems from the fact that whereas the individuation of content essentially depends on three different dimensions, the NCA is only committed to individuating mental states according to their narrow causal pattern. The other two dimensions, ideological and referential (or, truth conditional) similarity are to be amputated. First, these last two are unnecessary and therefore contribute nothing to the explanation and prediction of behavior. Second, by getting rid of them, context-sensitivity is eliminated. That is because, as in every multi-dimensional similarity judgment, it is the context that decides which dimension is to be emphasized in deciding whether a given state in a particular situation counts as the belief that P. Sometimes referential similarity will count more, sometimes ideological similarity, or simply causal pattern similarity depending on the demands of the particular situation in which the question arises.

Stich puts the greatest emphasis on the problems created by the *ideological similarity dimension*. This is what he calls the *holism problem* in the folk conception of belief. In order to bring out the problem vividly, let’s focus on his most celebrated thought experiment: the case of Mrs. T. Mrs. T is an elderly woman who suffers from a progressive loss of memory. At the end, she does not “know” what an assassination is, what dying is, who McKinley was, etc. Nonetheless, she appears to remember/believe that McKinley was assassinated, because that is what she persistently says when asked “What happened to McKinley?” According to Stich, the folk psychology’s clear verdict is that she does not believe that McKinley was assassinated. Stich’s diagnosis is that when she ceased to have a certain set of relevant beliefs, she ceased to believe that McKinley was assassinated, despite the fact that she appears to respond correctly to the

question. This, Stich says, shows that folk conception of belief attribution attends to the doxastic background of an agent. From this he seems to infer that the type-identity of someone's belief is partly constituted by what other actual beliefs the individual happens to have. This is the notorious problem of content holism, according to Stich.

On the other hand, an individuating scheme based on a NCA, he claims, is and ought to be nonsensitive to the actual doxastic surrounding of a mental state it individuates. That is because the NCA taxonomizes her state underlying her utterance on the basis of its *potential* (narrow) causal interactions. Thus STM is able to account for her ability to infer, for instance, "McKinley was buried in Ohio" from her "acknowledgment" of "McKinley was assassinated" and "if McKinley was assassinated then he is buried in Ohio." So whereas content psychologies miss such important generalizations as those that cover Mrs. T STM theories will be able to take such agents under their scope (for more on this, see below).

With respect to the *reference similarity dimension*, the situation is similar. For instance, there are many cases in which the causal history of certain subjects' use of some terms are so tangled that there is simply no saying what they refer to. And this makes it difficult to find comfortable characterizations of the content of the beliefs the subjects express using those terms. Hence, the difficulty in subsuming them under content generalizations. But for STM theories there is no such difficulty. The underlying states that cause those subjects to utter a sentence whose truth conditions are moot and the ones that we utter can be categorized as tokens of the same B-state mapped to the same syntactic object on the basis of their narrow causal potential. Similarly for those utterances whose referential properties are sensitive to the linguistic community of the subjects. The NCA is insensitive to such links outside of the agents' skin, and in being so, it is perfectly apt to capturing the regularities that ought to be under the proper domain of cognitive psychological theories.

The case with the *causal pattern similarity* is different. Those who are sufficiently causal pattern dissimilar to us will be beyond the reach of content generalizations because it will often be very difficult to find content sentences to characterize what intentional states they are in. For an STM style theory, however, this is no problem, or so claims Stich. All the STM theorist needs is a different set of generalizations, if the case at hand warrants it. Since the type identity of mental sentences depends substantially on which generalizations cover it, and since the causal patterns are different, different generalizations will be needed. Therefore, the abstract syntactic objects that the brain states are mapped onto will not be comparable to those of us. So there is a certain parallel here between the content and syntactic theories. However, unlike content theories, STM theories have no difficulty in stating these new generalizations, since all it has to be done is to postulate new syntactic objects in terms of which new generalizations are stated without any regard to what other *actual* brain states are mapped to which syntactic ob-

jects and to what referential connections obtain between the states and the objects outside the skin of the subject.

So, Stich claims, only the NCA, and hence STM, is what we ought to have for a successful and scientifically respectable cognitive psychology, if we don't want to miss important generalizations and cripple with a pervasive vagueness. If content based psychologies as analyzed by Stich were to be on a par with STM with respect to their virtues and vices, there would be very little point on the part of Stich to insist that we have to couch our generalization syntactically and forget about content. But, according to Stich, content and syntactic taxonomies radically diverge, so we have to choose. And it is the STM we have to choose because it is explanatorily superior to content psychologies: it has all the virtues of content psychologies and none of their aforementioned limitations. Let us now see whether Stich is right about his claim.

5.2 The Parallel Disadvantages of STM-Style Theories

In this subsection, I will take up all the three dimensions one by one, and show that either STM is no better off with respect to each of them or the criticisms leveled against content-based psychologies are unwarranted. However, Stich's emphasis is understandably on the "holistic" nature of beliefs, and so will be mine. I will start with the discussion of the referential dimension and argue that Stich is wrong in assuming without any argument that a scientifically respectable intentional psychology *must* be committed to broad content. I will then take up the doxastic similarity dimension. In fact, Stich's most important criticisms concern this dimension. I will show that Stich's STM has exactly parallel problems. Finally, I will briefly discuss the causal pattern similarity, and argue that this dimension does not pose any special problems for an intentional psychology, and STM has parallel problems anyway.

5.2.1 The Referential Dimension and Narrow Content

Stich rightly argues that commonsense individuation of propositional attitudes attends to reference or truth conditions. In other words, the folk individuate beliefs (*inter alia*) according to their broad content. This is one of the three dimensions mentioned above. In arguing against the prospects of incorporating folk notion of belief (and other propositional attitudes) into a respectable science of psychology, it is this notion of belief with broad content that Stich has in mind. Stich, of course, is entirely right in thinking that broad content does not supervene on what is inside the head, and therefore, you can, and sometimes do, have causal pattern identity without reference identity, and this *seems* to show that the broad taxonomy of attitudes plays

no essential role in psychological explanation. Hence, one of Stich's reasons for his case against belief: as far as commonsense individuation is committed to a broad taxonomy, it is likely to be eliminated.¹²

Stich does not consider the possibility of a certain form of revisionism *vis-à-vis* folk psychology. Granted that folk individuate beliefs broadly and what a scientific psychology needs is a taxonomic scheme that supervenes only on what is inside the skin, it simply does not follow that the notion of content has no place in cognitive science, unless there is no other notion of content but a broad one. But, many people in the field think that there is a notion of content that is narrow, and thus supervenes only on what is inside the head. Fodor, for instance, thinks that narrow contents are functions from contexts to truth conditions.¹³ These functions are implemented in the brain. Two thoughts are identical in their narrow content if: for all contexts *C*, one has the truth conditions *TC* in *C* iff the other has *TC* in *C*. In other words, context plus narrow content gives broad content. Intuitively put, narrow content is what you get when you subtract the contribution of the context from broad content.

There are many complicated and subtle issues about the notion of narrow content. I will not go into discussing them here. The important point I want to make against Stich is to distinguish between two strategies in arguing for the prospects of having a scientifically respectable *intentional psychology*. One is simply to adopt all the intentional notions of folk psychology with their taxonomic scheme without any filtering attempt to press them into scientific use. Fodor calls this strategy "die-hard".¹⁴ The other is to go revisionist. This strategy is ready to give up many things that the folk assume in so far as what remains uses an apparatus that is *essentially* and in some form recognizably *intentional*. Unfortunately, Stich does not distinguish between these two roads. Throughout his discussion what he seems to have in mind is the die-hard strategy.¹⁵

¹² The thought experiment based on twin-earth cases was due to Putnam (1975). Stich elaborated on this in his (1978) where he gave his celebrated "autonomy argument" for a content-free psychology. Fodor's methodological solipsism (1980a) was also taken by many as an argument for the elimination of semantics in psychology.

¹³ Fodor in his (1994) has changed his views about narrow content. He thinks that we can make do without any appeal to narrow content, so he rejects the notion. My reference is to earlier Fodor (1987, 1990, 1991).

¹⁴ Fodor (1989), p.175. Fodor in fact makes it quite clear that he is even ready to give up a propositional-attitude psychology as far as the psychology that scientifically survives is still intentional.

¹⁵ See, for instance, pp. 221–28 (1983) for a confusion of reductionism (hence, vindication of intentional psychology) with eliminativism. For a particularly striking example of his attack on a die-hard conservatism, see his modularity requirement on the part of folk notion of belief, p.237ff. (1983), and Ramsey, Stich, Garon (1990).

The notion of narrow content is a perfectly intentional notion: its identity conditions are given in terms of the identity of *truth conditions* across *contexts*. Put differently, narrow content is semantically evaluable given a context. So, as observed by Devitt (1990), a revisionist that opts for the notion of narrow content is not necessarily jeopardizing the prospects of scientifically vindicating folk psychology.

In the absence of an argument on the part of Stich to show that there can be no such thing as narrow content, I will simply drop the issue of reference similarity dimension for what follows as a red-herring. As we will see shortly, however, this will make little difference for what I am going to say against Stich. Stich, in fact, himself seems to have dropped the issue in criticizing Fodor's notion of narrow content in his (1991) article. He seems to grant the possibility of an intentional narrow psychology that is not committed to reference or truth-conditions. However, he has the same line of criticism against narrow content anyway: since narrow content, in some sense, is a construct out of broad content, the former has all the limitations that the latter has, except, of course, those that stem from the dimension of reference (or, truth conditions) similarity. This is not surprising, because, as I said, Stich views the holistic nature of intentional states (narrow or wide) as the most important argument against any sort of intentional psychology.

So we are left with the other two dimensions and their alleged trouble causing features for content based psychologies. Let us now see whether a taxonomy based on NCA has any superiority with respect to other two dimensions over content taxonomies.

5.2.2 Doxastic Similarity Dimension and the Holism Problem

Most of Stich's arguments for his case against belief turns on the "holistic" nature of commonsense individuation of beliefs. It is unfortunate that he does not elaborate on what, exactly, holistic nature of content comes down to. Much of what he has to say on the matter is provided through a handful of examples like the case of Mrs. T. This vagueness is typical on those who think that semantic content is holistic.

In what follows, I will first try to explain as clearly as possible the sense in which Stich thinks that commonsense individuation of beliefs is holistic. My discussion of Stich on holism of folk psychology will show that he is obscure and vague about what he thinks the "holism problem" is. I will then indicate exactly why Stich thinks that the holistic nature of belief is a problem for content based-psychologies, and why he thinks that an STM-style theory is totally free of it. We need to be as precise as possible about this, because my argument against Stich depends on his own premises.

5.2.2.1 Stich's "Holism"

According to Stich, the identity of a particular belief, say, the belief that McKinley was assassinated, depends on what other actual beliefs a person happens to have. The doxastic surround of the belief that McKinley was assassinated is constitutive of its content. But Stich does not say what this doxastic surround is, how it is determined, nor how it is supposed to be constitutive of a given content. For instance, at any moment, a person who believes that McKinley was assassinated has also a very large stock of other beliefs. Do all of them contribute to the content of the belief in question, or only some of it? From the way Stich writes and uses the expression 'doxastic surround of a belief' and its cognates, it seems that only some portion of a person's entire belief system, he thinks, is relevant to the determination of the content. Unfortunately, he gives almost no clue about how big the portion is supposed to be. At any rate, we have no idea about what, exactly, makes commonsense content individuation *holistic*. Where does the holism come from? The paradoxical aspect of Stich's discussion lies in here. Stich claims that commonsense taxonomy of beliefs is holistic, but all he may be said to show is that, at most, the identity of the belief that P is (partly) determined by *some* other beliefs a person actually has. If we reflect on his examples, this is obvious.

The case of Mrs. T is typical. She ceases to have many beliefs among which, for instance, is the belief that if someone is assassinated then she is dead. In fact, she no longer "knows" what dying is, what a presidency is, what presidents do, etc. In all of Stich's similar examples, the beliefs that make up the doxastic surround of a particular belief have rather certain "direct relations" to the belief. They are not only semantically close, but also, in some loose sense (to be discussed later), "conceptually" tied to the belief. In all such examples, the fact we are invited to observe is that when someone ceases to have *those* kinds of beliefs then someone ceases to have a particular belief.

When Stich presents his own descriptive analysis of what a content taxonomy comes to, he is equally unhelpful. Again, he does not give any clue about what portion of one's belief system, or how much of it, is relevant in determining any particular content. Just for the same reason, he does not discuss what, if only some portion is relevant, determines its relevancy. But he concludes that belief individuation is holistic!

I think, Stich is crucially vague and not particularly careful in his discussion. If his claim is that the content of a particular belief is (partly) determined by the set of *all* beliefs one has, which I take it what holism at its extreme comes down to, then he has not provided any single reason, let alone a relatively elaborate argument, for his claim. On the other hand, if his claim is that only *some* beliefs determine content, as he seems to intend, then the identity conditions for belief are not holistic.

I think that part of the reason why Stich is so vague and careless is that he does not care about this distinction, some or all. According to Stich, it seems, the very fact that the content of a belief depends for its identity on at least some other actual beliefs the agent has is enough to make serious trouble for any psychology that hopes to essentially advert to content in its generalizations. For one thing, given Quine's influence on him, Stich clearly thinks that the distinction between those beliefs that determine content and those that don't can be anything but sharp and principled. If so, content can at best be a matter of degree. And this is enough to make trouble: a content psychology is possible, at best, for those who are doxastically similar. But even for such a psychology, vagueness will still continue to be a serious problem, since it is almost certain that doxastic similarity never actually achieves doxastic identity among people. Whatever the case is with Stich's analysis of belief, however, he clearly thinks that his alternative paradigm of doing psychology, STM, does not have any such problem.

Let us see why and how Stich thinks that the framework provided by STM has no such "holistic" problem. Here is a typical remark by Stich:¹⁶

In chapter 7, section 3 our focus was on ideological similarity, and the persistent problem was that as subjects became increasingly ideologically distant from ourselves, we lost our folk psychological grip on how to characterize their beliefs. For a syntactic theory, however, ideological similarity poses no problem, since the characterization of a B-state does not depend on the other B-states that the subject happens to have. A B-state will count as a token of a wff if its potential causal links fit the pattern detailed in the theorist's generalizations, regardless of the further B-states the subject may have or lack. (1983:158)

Stich, then, goes on to clarify how this can be so by working on the example provided by Mrs. T:¹⁷

If we assume that before the onset of her disease the B-state which commonly caused her to say "McKinley was assassinated" obeyed generalizations like (4)-(6), then if the illness simply destroys B-states... without affecting the causal potential of the tokens which remain, the very same generalizations will be true of her after the illness has be-

¹⁶ For some others, see pp.53-60 and pp.137-44, Stich (1983).

¹⁷ For a similar and more striking discussion of the commitments of NCA of typing where Stich goes through a similar example, see pp.53-4 (1983). The generalizations (4)-(6) Stich mentions here are all what I will call below to be L-generalizations. They advert to the logical form of the sentences, hence are blind to the non-logical primitives the theorist postulates.

come quite severe. In chapter 7 we imagined a little experiment in which, shortly before her death, we tell Mrs. T, “If McKinley was assassinated, then he is buried in Ohio,” and she replies, “Well, then, he is buried in Ohio.” This is readily explainable by (5) [the syntactic version of psychologized Modus Ponens]... So if the generalization is there, it can be captured by a syntactic theory. But as we saw, there is no comfortable way to capture this generalization in the language of folk psychology... Thus a cognitive science that adopts the STM paradigm can aspire to broadly applicable developmental, clinical, and comparative theories, all of which are problematic for a content-based theory because of the constraints of ideological similarity. (1983: 158-9)

Is it true that ideological similarity poses no parallel problems for STM-style theories? I think not. It is time to see why.

5.2.2.2 Holism and STM

Here is the structure of the argument for my claim that STM, contrary to Stich’s advertisement, has exactly the parallel problems.

- (1) The STM framework is committed to the NCA of type individuating B-states *qua* mapped to particular syntactic objects like, say, ‘Fa’, through the generalizations that cover them.
- (2) The NCA is capable of individuating such states *only if* it has enough generalizations of a certain sort, which I will call, S-generalizations.
- (3) If STM has S-generalizations among its stock of generalizations then it has all the parallel problems that Stich complains about content-based psychologies regarding the dimension of ideological similarity.

In the remainder of this section, I will make this argument stick. I take it that (1) is common ground (see §§ 3–4). Let me first argue for premise (2).

All the T-T generalizations Stich ever considers, by way of giving examples or otherwise, have rather a certain character: They all quantify over particular syntactic objects, i.e., they all use meta-variables to refer to classes of actually specified sentences that have a certain common “logical form”. Even from his passage above, it is apparent that when he talks about the causal interaction of the token that underlies Mrs. T’s utterance of ‘McKinley was assassinated’, the generalizations Stich has in mind are of this kind. Let me call the generalizations that quantify over *particular* brain sentences “L-generalizations”, since these apply to any sentences that have

a certain “logical” form. L-generalizations are all blind to the primitive non-logical vocabulary that the STM-theorist specifies.

It should be obvious that if all the T-T generalizations that go into the specification of the causal role F in the individuation of $B(i)$ for any i (see above) are of this type, i.e., if they are all L-generalizations, then there cannot be a *unique* causal role for each particular $B(i)$, which means that there can be no type individuation of B-states with particular syntactic objects mapped onto them. Here is why: with only L-generalizations in force, *any* sentence token has *potential* inferential (causal) connections to *any* other one. Put differently, since, on Stich’s own admission, the generalizations in the theory detail not only the actual but also the potential causal interactions of any particular B-state, and since any sentence token can potentially be “inferred” from any other (i.e., causally connected through L-generalizations to any other), L-generalizations all by themselves cannot type individuate particular B-states.¹⁸ All they can specify is at most the “logical” form or syntactic type of sentence tokens. As we will see in the next section, this situation does not change even when we add the I-T and T-O generalizations to L-generalizations: together they are still incapable of providing unique causal roles for particular B-states.¹⁹ For one thing, as I will argue in the next section, there can be no such (narrow) I-T/T-O generalizations. But, for our purposes here, more importantly, even if there are such generalizations they can at most help to identify a very small subset of particular B-states whose character is rather “observational”. However, Stich himself is pessimistic of there being any such subset (see below). My point is that S-generalizations are necessary (not sufficient) for type individuating at least some B-states, and this will do for premise (2).

What is needed, of course, is a different kind of T-T generalizations in addition to L-generalizations, T-T generalizations that are not blind, so to speak, to the primitive non-logical vocabulary of the STM-theorist, generalizations that detail (part of) the causal role that is unique to, say, the B-state mapped to ‘ Fa ’. It is obvious that such “low-level” generalizations will typically be the syntactic parallels of such “content generalizations” (*C-generalizations*) as²⁰

¹⁸ It is ironical, and in fact a bit puzzling, that Stich himself makes the parallel point in criticizing content functionalism: “There are literally infinitely many inferential paths leading both to and from every belief” (1983, p.24). His point is that since every particular belief is potentially connected to every other, the generalizations detailing this potential will not be able to define beliefs with particular content.

¹⁹ In fact, the situation is even more complicated given that there is already a build-in vagueness in the “syntactic” individuation of particular B-states: for Stich a sentence token to count as of a particular type, it must satisfy a substantial number of generalizations. Stich seems to propose a cluster theory of type individuating sentence tokens, and this, as Stich himself admits, brings with itself a certain amount of vagueness. See below.

²⁰ A parallel distinction is drawn by Loar (1982a) between “L-constraints” and “M-constraints”.

- (i) For all subjects S and for all x, if S comes to believe that x is a cow, then S will typically come to believe that x is an animal,
- (ii) For all subjects S and for all x, if S comes to believe that x is a bachelor, then S will typically come to believe that x is unmarried,

and so on.²¹ Let me call the *syntactic parallel* of this kind of C-generalizations “S-generalizations”. Stich is committed to such generalizations, otherwise there is no individuation of particular B-states. Hence, premise (2).

Let’s now take up premise (3): If Stich is committed to S-generalizations, then his STM framework has exactly the same “holism” problem that he claims plague content-based psychologies. There are different ways of showing this, but at the end they all come to the same thing. Let me begin with the obvious version.

S-generalizations are *low-level* generalizations. What makes them low level is the following fact. Subjects that are covered by such generalizations are also covered by L-generalizations if the subjects have certain *actual* B-states. For instance,²²

if S has the belief* that #for all x, if x is a cow, then x is an animal#,

and

S comes to believe* that #Samantha is a cow#,

then

S will typically come to have the belief* that #Samantha is an animal#.

What might license this inference* is, of course, the existence of high-level L-generalizations that Stich (*mutatis mutandis*) specifies among his examples:

²¹ These are supposed to be “*ceteris paribus*” generalizations. I’ll generally ignore this in what follows.

²² In what follows, in order to avoid long and cumbersome ways of expressing the same thing, I will simply adopt the following convention: I will mark an intentional expression with a ‘*’ to express whatever its syntactic parallel may be. Also, I will hedge a content sentence with ‘#’s in order to indicate that I intend its syntactic parallel, i.e., whatever syntactic object or sentence might go in its stead.

(5) For all subjects S, and all wffs A and B, if S has a B-state mapped to A → B and if S comes to have a B-state mapped to A, then S will come to have a B-state mapped to B.
(1983:155)

As we may recall, according to Stich the “holism” problem that plagues the content-based theories consists in the fact that the type identity of a particular belief (partly) depends on what other actual beliefs the subject has. And Stich thinks that this fact is the source of the problem. In contrast, he claims, the NCA of typing particular B-states has no such commitment to there being any actual B-states surrounding a particular B-state in terms of which its type-identity is determined.

But, if every subject who is covered by S-generalizations is also covered by the relevant L-generalizations, in the way I’ve just indicated, then the STM-theorist is committed to their being actual B-states for determining the type-identity of particular B-states, and thus committed to construct syntactic theories only for those who more or less share their doxastic* background. In other words, in the STM paradigm, the “syntactic” type identity of sentence tokens is, contrary to Stich’s advertisement, acutely sensitive to the *actual* particular B-states that surround them. This is a problem that is exactly parallel to what Stich calls the “holism” problem of belief individuation. And so, STM must incur all the parallel problems that Stich claims seriously bother content psychologies: Sharing a particular B-state can only be a matter of degree, therefore, those that are doxastically* dissimilar to us cannot be covered by STM-theories. What are we to do with the children, exotics, cognitively handicapped, higher animals, etc.? Furthermore, unless Stich can come up with a principled distinction between those B-states that contribute to the syntactic type identity of a sentence token and those that don’t, the vagueness that is already existent in the conditions that type identify sentence tokens will be greatly aggravated. Again, we have exactly the parallel problem here. If Stich is right in his criticism of content-based theories regarding “holism” problem, it is false that STM theories are any superior in just that respect.

However, one might object: It is not necessary that for any subject who is covered by S-generalizations is also covered by the relevant L-generalizations in the way I have just indicated, and therefore, it is not necessary for an STM theory to be committed to there being *actual* B-states, which a subject must have, for the type individuation of sentence tokens. It may be that the S-generalization (i) above may be true of a subject even though she may not have any *actual* belief* that #all cows are animals#. In such a case, the syntactic type identity of a sentence token may be given in terms of such *dispositions* as the likes of (i) and (ii) specify without

any recourse to high-level L-generalizations.²³ How does this evade the problem? Well, let me show that it doesn't.

Although I don't have to make my point in the way I will do, I think it is important to cast the issue in terms of that perspective. But nothing important will hang on it. STM has usually been brought up as a de-intentionalized version of a language of thought story, or CTM. We have seen that STM is not in fact committed to there being (semantic-free) sentences literally realized in the brain. But it may be taken in this way, and this is the assumption we are now operating under.

Anyone who is sympathetic to the computational paradigm must keep in mind that CTM is a "rules and representations" framework: any relatively higher level mental processes consist in transformation of syntactically structured representations according to rules that are causally sensitive only to the formal properties of representations over which they are defined. In other words, the typical computational treatment of such inferences as expressed by (i) or (ii) will take the form of applying some relatively high level rule like Modus Ponens to *actually tokened* complex sentences.

Of course, this is only one possible implementation story that can be given for such generalizations like (i) or (ii) at the computational level. Another possibility is that the rules that govern the inference from #x is a cow# to #x is an animal# is rather more specific and low-level, rather like the syntactic analogues of Carnap's "meaning postulates" implemented as rules.²⁴ But, either way, according to CTM, you need rules to manage inferential processes defined over data structures.²⁵

My point is simply this. On a computational paradigm S-generalizations can be cashed out *either* by postulating high-level laws and actual beliefs* upon which they operate *or* by postulating the syntactic analogues of meaning postulates in the form of low-level rules. So, once it is obvious that an STM-theorist is committed to such S-generalizations, the STM-theorist is no

²³ Devitt (1996) draws a similar distinction while criticizing meaning holism and preparing his way toward a defense of meaning localism. According to Devitt, the meaning of a (mental) word is (partly) constituted by its inferential properties, where this unpacks as: the meaning of a word is constituted by some of the inferential roles of the sentences that contain it. On Devitt's view, this is not to be confused by the following less plausible view: the meaning of the word is constituted by some of the beliefs that contain it. The former is what he calls the inferential version of meaning localism, the latter the belief version.

²⁴ There are many versions of this approach in AI. Frames, scripts, etc. are all versions of the same underlying idea. The tradition of "semantic representation" in linguistics again relies on the idea that lexical items can be semantically decomposed.

²⁵ Rules may or may not be explicitly represented. CTM is neutral on this. However, given that the rules that implement S-generalizations reflect important pieces of "semantic knowledge" they are unlikely to be hard-wired.

longer in the position that Stich claims is free of the problems confronting the content psychologist. Let me illustrate.

As we may remember, Stich claims that unlike content psychologies, the STM framework is capable of covering Mrs. T's mental states in its generalizations. That is because, he says, according to STM, the type identity of the state underlying her utterance "McKinley was assassinated" does not depend on further actual doxastic* states she has. In fact, as the example is constructed, she has almost none. The type individuation of Mrs. T's state proceeds according to its inferential* *potential*, not according to what it *actually* inferentially* interacts with. So far so good. For instance, all the L-generalizations that cover her state detail just this potential. But, of course, with only L-generalizations, the STM-theorist cannot individuate the state. It is obvious that in addition to L-generalizations, the STM-theorist needs some such S-generalizations as

(iii) For all subjects S, if S comes to believe* that #someone is assassinated#, then S will typically come to believe* that #someone is dead#

in order to type individuate the state underlying Mrs. T's utterance of "McKinley was assassinated". But, of course, we see that it is precisely this kind of generalizations that become inapplicable to Mrs. T when we come to see that she ceases to have many relevant beliefs. This can easily be explained on the version of the computational story that derives the S-generalizations from actual beliefs* and high-level L-generalizations. But Stich would probably insist that this is the wrong version. Well, then, let us look at the other version where S-generalizations are implemented as specific "dedicated" rules rather like the syntactic analogues of "meaning postulates".

The question now is whether there are any such rules intact in Mrs. T's case. As we may remember, it becomes apparent *under questioning* that she does not "know" whether an assassinated person is dead, what dying is, who McKinley was, etc. When she is asked whether McKinley was dead, she answers "I don't know". What better evidence can there be for the fact that the above S-generalization is broken? In the case of S-generalizations, appeal to potential causal profile doesn't even begin to help since it is precisely this potential that is lost in her case. But then, if such generalizations do no longer cover the mental states of people like Mrs. T, of course, we can't tell the computational story along the lines we have been assuming given that the other version is out. But either way, the important point is that the S-generalizations do simply not hold in Mrs. T's case. If so, however, the STM theorist is in exactly the same boat as the content psychologist: there is simply no saying what "syntactic" state Mrs. T is in, since the STM-theorist is no longer able to type individuate her state.

The same is true, similarly, for people who are doxastically* dissimilar to us like children, exotics, cognitively handicapped, higher animals, etc. In so far as the S-generalizations that are true of them are not available or non-existent, there is no type individuation of their syntactic mental states, hence they are beyond the reach of STM-theories.

So here is the score. Contrary to Stich's claim and advertisement, because of their commitment to NCA, STM theories are committed to the type individuation of particular B-states either (depending on the computational story preferred) according to what other actual B-states the subjects have, or according to what S-generalizations are true of them. The first option makes STM equally sensitive to the *actual* doxastic background of subjects. The second option restricts the scope of STM-theories to those for whom S-generalizations exist, or are specifiable, thus, again, to those who are doxastically/dispositionally similar. But the consequences of both options are just the same for the prospects of STM if the prospects of content psychologies are as Stich claims them to be.

Before ending this section, we have to take care of one more thing. Stich's analysis of belief includes, remember, one more dimension in the similarity matrix that goes into the commonsense individuation of beliefs: the causal pattern similarity. Let us now see whether this dimension poses any special problems for content-based psychologies in a way STM-theories are free from.

5.2.3 Dimension of Narrow Causal Pattern Similarity

According to Stich, when certain subjects are sufficiently dissimilar to us with respect to the causal pattern dimension (call it cp-dissimilar), there is no saying what they believe. But since STM style theories do not traffic in intentional terminology, they can come up with different generalizations, hence with different syntactic objects, that detail these subjects' functional/causal organization. Cp-dissimilarities can be attributed to three different sources:

- [1] dissimilarities that stem from different doxastic background (as in the case of exotics, etc.),
- [2] dissimilarities that stem from different or broken down ratiocinative mechanisms (as in cognitively handicapped people — different L-generalizations), and
- [3] dissimilarities that stem from perceptual or behavioral differences (as in the case of perceptually or behaviorally handicapped people — different I-T or T-O generalizations).

With respect to [1], Stich is no better off than the content-based psychologists. Here is why. First, Stich thinks that the cp-similarity dimension is distinct from the other two dimensions. But this can't be quite right. In particular, since Stich is committed to S-generalizations,

the cp-similarity cannot be an entirely separate dimension different from doxastic similarity dimension. S-generalizations are supposed to detail nomological connections among beliefs* with particular #content#, hence they are rather specific and dedicated generalizations that hold among doxastically* similar subjects. So, in a certain sense, the doxastic similarity issue turns out to be an aspect of cp-similarity issue. They collapse into one dimension. To put it differently, the class of subjects who are sufficiently cp-dissimilar to us include those who, Stich calls, are doxastically dissimilar. So if exotics and the like pose a problem for intentional psychology, by parity of the cases, they should pose a parallel problem for STM theories. If, on the other hand, the STM-theorist can come up with different S-generalizations for doxastically dissimilar subjects, as Stich might claim, why can a content psychologist not do the same? In fact, given that the reference dimension is put aside, there is every reason to believe that this is in fact what they do. Anthropologists routinely describe in intentional terms the doxastic network of such culturally remote groups: they seem to detail, in other words, the network of C-generalizations that hold for such groups.

Stich similarly claims that where differences stem from [2] and [3], the STM theorist can always state new generalizations. But again why can't the content psychologist do the same, if this is the proper way to go?

Stich would probably say that (broad) content attribution is necessarily parochial, whereas attribution of syntactic objects on the basis of new generalizations is not. But we are assuming that a scientific intentional psychology will not advert to broad content but to a narrow one. Also, given that the cases are exactly parallel, it should be the case that either the two types of theories both can't cover causally dissimilar subjects or they both can.

I think, Stich is forcing us again to be die-hard conservatives in defending a content psychology, and he's been wrong on this. The essential question is whether the dissimilar subjects are intentional agents, and not whether *we can attribute* broad content to them. In cases like [2] and [3], as long as the subjects are intentional, much of what matters will be whether *identity* or *difference* of the narrow content of their mental state is defined. And, as long as the answer to that question is positive, the content psychologist can perfectly well quantify over the content of their states.²⁶ And there is no reason to think that the answer is negative. Or if it is, then, similarly, specifying the identity or difference for abstract syntactic objects Stich wants to operate with is equally problematic in the STM paradigm.

My present point however is that, given the cases are exactly parallel, Stich needs an argument for his claim that STM theories are superior in the aforementioned respects to content

²⁶ Cf. Fodor's similar remarks in his (1987), pp.161-3.

based theories. And he doesn't have one. I suspect that the reason for that is his confusion about what STM theories are or are not committed to.

In fact, as we will see in the last section of this paper, an intentional psychology has even an important advantage over STM theories in their ability to discover and state new set of generalizations for those that are sufficiently cp-dissimilar to us. I will argue that especially in the context of discovery, an STM-theorist has no chance but to become an intentional psychologist.

In conclusion, I claim to have established in this section the following conditional claim: *if content based psychologies have the problems Stich enumerates then STM-theories have exactly the same problems.* So Stich loses all his motivation to promote STM as a superior paradigm of doing scientifically respectable psychology.

Now it is time to see whether the NCA can type-individuate the particular syntactic states at all, as Stich assumes. In the following section, as advertised, I will argue directly against STM by showing that it can't.

6. The NCA and the Type Individuation of Brain Sentences

[[Dear reader: The following subsections §§ 6.1 & 6.2 were written after a considerable period of time had passed since the first draft of this paper was written. I hope the difference in style and some (mostly terminological) discrepancies between the material in this section and the others aren't big enough to cause serious problems in following the thread of the argument in the paper. I also hope that some overlap and repetitions would be excused.]]

As we may remember, Stich had raised a problem when he was discussing Fodorian mental sentence theories:

for a Fodor-style account of belief sentences to hang together, we must have some workable notion of what it is for two distinct people, speaking different languages, to have in their heads distinct tokens of the same sentence type. (1983:43-4)

Among the three possible accounts he offered, he rejects two, the Quasi-Physical Account and the Semantic Account (SA), because he thinks that either they can't do what they are supposed to do or they are ill-suited for serious scientific purposes. What remains is the NCA. Since Stich intends to take his STM rather in the spirit of mental sentence theories, he makes the NCA the basis of his allegedly superior STM in type identifying the brain sentence tokens.

In this section, I will first show what was supposed to be wrong with narrow content functionalism. I will then argue that Stich's STM is identical to narrow content functionalism, and as such suffers exactly from the same problems: namely, NCA-cum-STM either doesn't have the resources to type-individuate brain states with particular syntactic objects across different organisms in a way appropriate for the purposes of scientific psychology, or else suffers from the same radically interactable "holism problem" that Stich says plagues content based psychologies. We will see that these two problems are at bottom different aspect of a single problem, so at the end, it doesn't really matter which horn of the dilemma Stich's theory will get stuck to.

6.1 Troubles with Content Functionalism

Let us begin by reminding ourselves what was supposed to be wrong with functionalism in fixing mental content. Fodor has been the most vocal person in the last ten years or so in his attack on semantic functionalism. Partly because of this, I will loosely take Fodor's attack on semantic functionalism to guide us through the steps of the argument that will show what the troubles are with content functionalism. Where necessary, I will clarify, supplement and correct Fodor. In a way, here I am siding with Fodor about the the place of functionalism in fixing the type-identity of particular Mentalese symbols against Stich. I think that Stich has misunderstood the role of functionalism in cognitive science.

Fodor's sustained attack on content functionalism has changed in emphasis and argument over the years, and there are many obscurities in the details. In the beginning, the emphasis was on the holistic consequences of content functionalism, which were thought to be destructive of a scientifically respectable intentional psychology.²⁷ Later on, he argued that functional role semantics either violates semantic compositionality or else is committed to there being a principled analytic/synthetic distinction, and neither option was thought to be palatable.²⁸ Although his earlier and later criticisms are very intimately connected, I will concentrate in what follows only on his earlier views, since they are the ones that are relevant to criticism of STM.

According to narrow content functionalism, the content of a Mentalese symbol token is (metaphysically) determined or constituted (partly) by *some* of the inferential relations it has.²⁹

²⁷ See, for instance, Fodor (1985, 1987:Ch.3).

²⁸ See Fodor and Lepore (1991, 1992:Ch.6).

²⁹ I want to leave aside the issue of whether the content of a token that is to be fixed by its functional role is broad or narrow. It is in a way natural to view it as narrow since we are dealing narrow functional role, and since, on any reasonable theory of semantics, referential links to reality must somehow be made part of the overall theory explaining meanings/contents. But this won't be important in what follows. Suffice it to say that what is being fixed is thought

It is crucially important to be clear about the notion of inferential relations of a token. I want to make six points (in no particular order) to clarify the notion.

First, a content functionalist aspiring to naturalize intentionality, or at least conceiving functionalism as part and parcel of the project of naturalizing semantics, cannot, unquestion-beggingly, appeal to a token's "inferential relations" as such, since the very notion of inference is an intentional one. As noted by Fodor,³⁰ the usual solution to this problem has been to combine functional role semantics with LOTH or computationalism in general. The idea is that inferential relations are to be cashed out by computational/functional/causal relations among non-intentionally characterized symbol tokens.³¹ This point is very important. Indeed this is the first point where content functionalism and the NCA of typing tokens comes into contact, since a content functionalist is now in need of non-intentionally individuating symbols so that she can assign semantic content to them on the basis of certain of their computational relations. I will come back to this point below.

Second, as I pointed out previously, a symbol token, in a certain obvious sense, has indefinitely many *potential* inferential relations.³² For instance, #Fa#³³ can be inferentially connected to *any other* token provided that the agent has the further appropriate tokens and the "syntactic" (or, proof-theoretic) version of some of the basic *logical* generalizations (L-generalizations) like Modus Ponens, Conjunction Elimination, etc., in her computational repertoire. We called them L-generalizations. So, for example, #Fa# is inferentially connected to #Y #, through Modus Ponens* and the expression #Fa Y # for any predicate Y and any constant . L-generalizations achieve generality by quantifying over *specific* symbols. Modus Ponens, for instance, says that given *any* conditional and its antecedent, its consequent may be in-

to be semantic, as indeed intended by many two-factor semanticists.

³⁰ Fodor (1987:75–6), Fodor and Lepore (1991:336–7, 1992:179).

³¹ Although this is the natural and usual solution, I should note that a LOT framework is by no means essential in the solution of the problem. Loar (1982a) is a good example of a functional role semantics that is explicitly not committal about LOT. See his discussion in (1982a:205–8) and (1982b). For what follows I will assume the LOT framework for a functional role semantics. This is necessary for my polemical purposes at any rate.

³² In what follows, I won't distinguish between subsentential and sentential symbols. Although 'inference' is most naturally applies at the sentential level, we may, for the sake of convenience, stipulate that the inferential relations of subsentential symbols ("words") are to be specified by sentences in which they occur. See below.

³³ As per previous paragraph in the text, in order to avoid long and cumbersome ways of expressing the same thing, I will adopt the following convention in what follows: I will hedge a content sentence or symbol with '# 's to indicate that I intend its computational/"syntactic" parallel, i.e., whatever "syntactic" object or sentence (general or specific) might go in its stead. Also, although I will be relaxed about it in what follows, when it seems to matter, I will mark an intentional expression with a '*' to express whatever its computational/"syntactic" parallel may be.

ferred. No particular conditional needs be mentioned. This makes it clear that the meaning-constitutive inferential relations of a token cannot be any such potential relations secured by L-generalizations. Put differently, since L-generalizations make any given token potentially connected to any other actual or possible tokens, Mentalese tokens cannot be type-individuated (intra or interpersonally) with only L-generalizations.³⁴ Rather, some of the individuating generalizations must be more restricted, immediate and specific. Above, we called such generalizations S-generalizations. Such generalizations must not quantify over specific symbols. Their quasi-canonical form can perhaps be given as such:

(C) For all *S*, if *S* comes to have #*F**x*# in her B-box, then, *ceteris paribus*, *S* will come to have #*G**x*# in her B-box.³⁵

Here are some putative examples, some of which are perhaps more likely to be true than others:

- (1) For all *S*, if *S* comes to believe* that #*x* is a bachelor#, then, *ceteris paribus*, *S* will come to believe* that #*x* is unmarried#;
- (2) For all *S*, if *S* comes to believe* that #*x* is assassinated#, then, *ceteris paribus*, *S* will come to believe* that #*x* is dead#;
- (3) For all *S*, if *S* comes to believe* that #*x* is a cat#, then, *ceteris paribus*, *S* will come to believe* that #*x* is an animal#;
- (4) For all *S*, if *S* comes to believe* that #*x* is a star#, then, *ceteris paribus*, *S* will come to believe* that #*x* is a celestial object#;
- (5) For all *S*, if *S* comes to believe* that #*x* is a tiger#, then, *ceteris paribus*, *S* will come to believe* that #*x* is dangerous#.

It is on the basis of such generalizations (in addition to L-generalizations) true of a token that its inferential relations are specified.

Third, although I specified the S-generalizations as quantifying over all subjects, *S*, it remains to be seen what the domain of ‘*S*’ will actually be for each generalization. At one ex-

³⁴ For further discussion of the nature and plausibility of L-generalizations as psychological generalizations, see Loar (1982a:71ff.). Loar calls them L-constraints, and adduce quite a number of interesting and plausible examples.

³⁵ “B-box” is meant to non-semantically capture whatever computational mechanisms underlie our belief forming and storing capacities. (C) as it stands is not in fact well-formulated: ‘*x*’ in ‘#*F**x*#’ has to be a meta-variable ranging over the “referring expressions” of *S*’s Mentalese. In what follows, I will ignore this and other technical complications involved in properly stating S-generalizations. This formulation and the following toy examples are meant to be understood at an intuitive level.

treme, the generalizations may be true of one organism; at the other, they may range over an entire population, or even species. True enough, content functionalism, if it is to vindicate intentional realism and a scientifically respectable intentional psychology, must not make the domain of such generalizations restricted to individuals or indeed to subcultures or perhaps even to cultures. Intentional psychology is meant to be the psychology of the folk. Hence content functionalists must mean their S-generalizations to cover at least the folk.

Fourth, as I mentioned in characterizing NCA above, the canonical way of specifying the narrow functional role of a token is by way of the Ramsey sentence of the generalizations covering that token. Informally, the (narrow) meaning of a token is to reduce to its functional role. This functional role is to be fixed on the basis of generalizations that cover it through the standard Ramsey-Lewis-Loar method. However, not all S-generalizations that cover a token might go into the Ramsey sentence that would determine its semantic content. Content functionalism typically assumes that only *some* of all available generalizations will go into determining content. Hence, we can talk of a “holist” functional role of a token with its Ramsey sentence containing *all* the available generalizations covering it, and we can also talk of a “localist” functional role with its Ramsey sentence containing only *some* of all the available generalizations covering the token.³⁶ The way I initially characterized content functionalism above is to be understood in this latter sense. (See below.)

Fifth, since, according to functionalism, the content of a token is meant to be determined by its localist functional role, and since distinct functional role means distinct content, and vice versa, it is assumed that the generalizations will fix *unique* roles for each intuitively distinct content. Thus it is assumed that there are *enough* generalizations in the relevant “localist” Ramsey sentence to fix a unique role for each distinct content.

Sixth, the meaning-constitutive generalizations must in some intuitive sense be *lawlike*. I am not sure what this means exactly. But at a minimum it is supposed to ensure that generalizations must go beyond describing statistical summaries of what has actually caused what, that they support counterfactuals in appropriate ways.³⁷ (I will come back to this point in a moment.)

Now given these clarifications and distinctions, we may characterize narrow content functionalism as follows. There is a theory T consisting of a set of S-generalizations³⁸ such that *via* Ramsefication they secure a *unique* functional role for any contentful token and this role *is*

³⁶ See Devitt (1996:43–7) for a somewhat similar line on how to properly characterize content functionalism.

³⁷ Cf. Stich (1983:27), Loar (1982a:44ff.).

³⁸ L-generalizations must also be part of T. I will assume that this is so in what follows and will not mention them again.

the (functional role — narrow) content/meaning of the token. The generalizations of **T** are true of all those with a reasonably common intentional psychology (more or less assumed to be the Folk with a certain common conceptual repertoire). Thus the lawlike generalizations are *interpersonally applicable*.

Let me emphasize again that the generalizations in **T** are not assumed to exhaust *all* the generalizations covering any particular token whose content is secured by **T**. In particular, there may be, in the first case, other equally interpersonally applicable (and, perhaps, lawlike — see below) generalizations covering all (or most, or some) tokens covered by **T**, but they need (or perhaps must) not be put in **T**. At least, this is a question left open. Intuitively, they may not be meaning-constitutive S-generalizations. Perhaps, (4)–(5) may be given as examples here. Call such interpersonally applicable S-generalizations not meant to be in **T** (hence not meaning-constitutive), *pluralist*.³⁹ In the second case, there may be many other S-generalizations only true of single individuals' tokens otherwise covered by **T**, an example of which may be:

(6) For all *S*, if *S* comes to believe* that #*x* is a bachelor#, then, *ceteris paribus*, *S* will come to believe* that #*x* is a neurotic#.

Suppose this is true only of me; so the domain of '*S*' is the unit set {Murat}. Intuitively, this would be the case, for instance, if I were to strongly believe that all bachelors are neurotic. Then it is true of only me that if *E* were the tokening of #*x* is a bachelor# in me, then *E* would tend to cause the tokening of #*x* is a neurotic# in me. Call such S-generalizations *singularist*. Then my #bachelor# token is covered both by (1) presumably in **T** and by (6) not in **T**.

Content functionalists' informal claim that the meaning of a token is constituted only by *some* of its inferential relations is to be therefore taken in the sense that only *some* interpersonally applicable S-generalizations covering a token can be meaning-constitutive, i.e., can go into **T** which is, *via* Ramsefication, the sole basis of securing a unique functional role for the token that is to be identified with its (narrow) meaning.

Now, what is wrong with content functionalism so characterized? Surprisingly, it is very hard to tell from the way people write on and complain about functionalism. Fodor is typical in this regard. Part of the reason, I think, is that Fodor ignores many of the above clarificatory points and distinctions in his discussion. But it is possible to discern his bottom line. He accuses the content functionalist of being holist.

Semantic holism is the doctrine that the meaning of a token is constituted by *all* of its inferential relations. But this statement carries with it all the obscurities of the original phrase 'in-

³⁹ Pluralist, because they cover more than a single individual. I am aware that the choice of

ferential relation.’ Given that the functional role of a token is fixed by the generalizations covering it, semantic holism, in the context of Fodor’s accusation, must be the doctrine that the content of a token is constituted by *all* the generalizations that cover it. But what is supposed to be wrong with this?

The official word for why holism is to be avoided is that it is destructive of intentional psychology: it is supposed to make meanings not publicly sharable. As can be seen from our discussion, a destructive sort of holism follows from functionalism only if by ‘all’ we mean *all the available* generalizations covering a token including not only the pluralist but also the singularist ones. It must be obvious that even if we decide to make the pluralist S-generalizations meaning-constitutive by incorporating them into T, a destructive sort of holism *still* doesn’t follow, not unless without some further assumptions. Call this sort of holist functionalism Non-Destructive (*ND-holism* for short). So Fodor’s worry about holistic functionalism must be a worry about a version of the theory that also takes the singularist S-generalizations into account in fixing a functional role, hence (narrow) meaning. Only then does functionalism become destructive.

But why take functionalism in this version? As Fodor put it once, in this form functionalism, as a naturalistic semantics, becomes suicidal: it ends up naturalizing meanings that no one seems to share. This wasn’t at all the notion of ‘meaning’ that functionalism had envisioned to naturalize to begin with. Indeed why take functionalism in this way?⁴⁰

Fodor seems to think that a localist or even a ND-holist functionalist is somehow forced to become a holist of a destructive sort. But what are the grounds for thinking that? Why is a localist functionalist forced to become a destructive sort of holist? Why can’t the content functionalist remain localist, or even a ND-holist, admitting as meaning-constitutive only those generalizations true of all those organisms with, intuitively, a reasonably common intentional psychology and discarding all those generalizations with intolerably narrow scope? Indeed, this way seems quite natural for a content functionalist who wants to avoid the destructive consequences of holism.

the label isn’t ideal, but the other choices looked to me at least equally unhappy.

⁴⁰ Especially when some of its most prominent advocates like Loar are so explicit about the version of their functionalism: “It is part of my project to vindicate interpersonal synonymy — that a given belief, classically individuated, can be predicated of different persons on objectively determinate grounds regardless of its evidential differences for them. So it appears we must find interpersonally ascribable generalizations about beliefs — that is, those that are counterfactually true of everyone’s beliefs, that collectively imply something unique about each belief individuated in the fine-grained way, and that belong to a commonsense theory...” (Loar 1982a:64).

Unfortunately, there is no clear answer to this question in Fodor. His official line is to be found in an argument he reconstructs on behalf of holists, which he calls the “Ur-Argument”:⁴¹

- (i) The semantic content of a Mentalese symbol is constituted by *some* of its inferential relations.
- (ii) No principled distinction can be drawn between those inferential relations that constitute the content of a symbol and those that don’t.
- (iii) Therefore, the content of a symbol is constituted by *all* its inferential relations.

(i) is apparently the position of localist functionalists. (ii) is supposed to have been established by Quine’s argument against drawing a principled distinction between analytic and synthetic statements. Hence (iii), which is the position of holists. As I said, in order for Fodor’s worry about holism to be warranted, (iii) must be read as referring to inferential relations captured by S-generalizations including the singularist ones. I will be charitable and assume that this is what Fodor has in mind.

But what does the analytic/synthetic (a/s) distinction have to do with functionalism? When reconstructed in our terminology, (ii) seems to say that there is no principled distinction between those generalizations covering a token that are in **T** (meaning-constitutive) from those that aren’t (non-meaning-constitutive). Is this claim the same with Quine’s claim on the untenability of the a/s distinction? And is it true?

The answer to the former question is, strictly speaking, negative, although as we will see in a moment the two claims are intimately related. Functionalists (local or holist) cannot appeal to intentional idioms like the a/s distinction in their attempt to explicate what meanings are. Being naturalists all there is in their disposal are causal/functional/computational relations. At best, they have to reconstruct the a/s distinction, if they deem it worth doing, in terms of such naturalistically acceptable relations as specified by causal/functional/computational generalizations.

But is it true that there is no principled distinction between those causal generalizations covering a token that are meaning-constitutive from those that aren’t? Here is one suggestion about how to draw such a distinction that seems natural given the dire consequences of destructive holism. A generalization is meaning-constitutive (hence in **T**) only if its scope is sufficiently large, intuitively, to accommodate all who have a common psychology.⁴² For instance, the very

⁴¹ Cf. Fodor (1987:60ff.), Fodor and Lepore (1992:23ff.) and Devitt (1996:10).

⁴² I assume that the circularity involved here in informally characterizing the idea can be avoided by a functionalist. The essential idea is to distinguish certain generalizations from others on the basis, *inter alia*, of their scope. See below.

reason why we shouldn't count (6) as meaning-constitutive is precisely because its scope is not large enough: it applies only to me. So the difference between (1) and (6) lies precisely in the fact that (6) is idiosyncratic in its application only to me, whereas (1) seems to be true of all those who have, intuitively, the concept of a bachelor. This move would establish a difference in such a way that avoids holism's destructive consequences. The suggestion tries to establish a difference between those generalizations that are in **T** from those that aren't in terms of their application scope or domain.

But, whatever other problems the proposal may have (and there are plenty), unless the localist functionalist moves into a ND-holist position, it falls short of its intended target, because the question then becomes: What decides on which ones go into **T** from among all those interpersonally applicable S-generalizations (i.e., generalizations with a sufficiently large scope)? Intuitively, suppose (1)–(5) are of equally large scope. What is to distinguish between them — if anything — as to which ones go into **T**? Is there, for instance, a difference between (1) and (2) on the one hand, and (4) and (5) on the other in that regard?

It is at this point one is tempted to appeal to a de-intentionalized version of a/s distinction. The problem here isn't that Quine was right and so there isn't a principled a/s distinction. For even if there were it wouldn't be available, as such, to a localist content functionalist on pain of violating the strictures of naturalism. So the functionalist must come up with a distinction, which would perhaps reconstruct the semantic a/s distinction in non-semantic (causal/computational) terms. What could the basis of such a distinction be?

There are at least two options available to the localist at this juncture. One is to admit that the job of finding such a naturalistically acceptable distinction is hopeless and move into a ND-holist position, to which I will come back below. The other is to try what seems to be an intuitively natural proposal: the difference between the S-generalizations that are in **T** and those that aren't (pluralist ones) is that whereas the former are truly lawlike, the latter are not. So far we haven't emphasized this aspect of functionalism, and assumed that all the generalizations we have talked about are equally lawlike. But perhaps this assumption isn't true. As I said, I am not sure what exactly lawlikeness comes down to in the case of S-generalizations, and so am not sure whether the following rough and ready characterization of the proposal can be further elaborated and successfully defended. I just want to put it on the table on behalf of the localist functionalist.

A generalization is lawlike only if it supports counterfactuals. (But not every counterfactual supporting generalization may be lawlike.) So, for instance, the generalization 'in *C*, (x)(*F*x → *G*x)' is lawlike only if it is true that in *C* if *this* were (had been) an *F*, it would be (would have been) *G*. Given some of the most popular accounts for evaluating counterfactuals, e.g. Lewis' (1973) and Stalnaker's (1968), the S-generalizations (1) through (6) and the likes (with

perhaps their differing domains and to various degrees), if true, turn out to be counterfactual supporting. This is hardly surprising. Such generalizations describe the *causal regularities* among state transitions of the systems under their domain. These regularities are described at the level of state types. Intuitively, a Ramsefied functional theory consisting of such generalizations describe the abstract functional/causal structure and organization of individual systems/organisms under its scope. If the theory applies only to me at a time, then it describes the causal *potential* of my state types at that time. So, for instance, suppose (6) is part of a functional theory of me, and only me, at t . Then it is counterfactually true of me at t that if I were to token #John is bachelor#, I would, *ceteris paribus*, tend to token #John is neurotic#.

For all that, on the other hand, (6) may not be lawlike if its application domain is only me and it is restricted to particular times. Unfortunately, there doesn't seem to be any consensus on a recipe of what needs to be added to a counterfactual supporting generalization to turn it into a law. But it is clear, at least intuitively, that laws are not the kind of things that are parameterized for particular times and populations who are relevantly homogeneous. If this doesn't seem intuitively clear with respect to laws or lawlikeness, we can perhaps introduce a term, "strongly counterfactual supporting" (*strongly-CS* in short), to capture this aspect of being relatively free from temporal and regional restrictions (in addition to whatever laws are beyond simply being counterfactual supportive), and define it thus:

a generalization is *strongly-CS* if and only if in all the nearby (otherwise nomologically) possible worlds, where the antecedent is true of the same population over a given period of time, its consequent is also true.⁴³

So for instance, it may be said that (1) and (2) are strongly-CS in this sense (or else, intuitively, there is no token with the contents [bachelor] and [assassinated], hence no concepts of bachelor and assassination), whereas (4) and (5) are not.⁴⁴ Take, for instance, (5). There seem to be indefinitely many nearby nomologically possible worlds in which no organisms tend to think that something is dangerous upon thinking that it is a tiger, and this seems to hold independently of

⁴³ This is weaker than the version where we don't even restrict the worlds to those that involve "the same population over a given period of time," but to any nomologically possible worlds in which cognitive organisms with sufficiently rich conceptual sophistication exist. Such a world wouldn't be nearby presumably. In the version given, we are supposed to imagine the same population having different developmental histories either in response to different nomologically possible environments, or in response to different cultural/cognitive pressures, or, typically, both. On the other hand, I am not sure to what extent we can keep the two versions distinct. But nothing will hang on this in what follows. The emphasis must be on the qualifier 'nearby.'

⁴⁴ I avoid (3), because I don't know where to include it.

whether in those worlds tigers, if they exist, are dangerous. Similarly for many others. In fact, it appears that we don't even have to look at other worlds. To make the point, the actual world will do. For instance, (4) is said to be false of ancient Greeks (it may even be false of some of our contemporaries) of this world: they apparently believed that stars were holes in the celestial spheres that the cosmic fire shows through (cf. Fodor 1987:88–9). That such generalizations are not strongly-CS in the sense defined, a localist functionalist might say, should hardly come as a surprise in fact. For, intuitively, their truth depends on what *actual beliefs** people have in common at a time, or what *common content*-specific inferences** they are *actually* disposed to engage at a time, and to a very large extent this doesn't seem to be a matter expressible in terms of nomological necessities.⁴⁵

We may put the problem slightly differently as follows. Among such generalizations, many do not seem to be necessary for a subject to have a particular concept*. For instance, for a subject to have tiger-thoughts*, it does not seem necessary that the generalization (5) should hold for her, even though (5) may hold for everybody else.⁴⁶ But this does not seem to be true of, at least, (1)–(2). There indeed seems to be a difference between (1)–(2) and (4)–(5). That, say, (1) should hold seems to be necessary for someone to have any bachelor-thought* at all. It seems that this is another way of saying that (1), if anything, is a strongly-CS generalization. Anyone for whom (1) is false *ipso facto* can't have bachelor-thoughts*, or so it seems. In other words, in all the nearby nomologically possible worlds in which bachelor-thoughts* are defined, anyone who comes to think* that #x is a bachelor# tends, *ceteris paribus*, to think* that #x is unmarried#.

Perhaps, this is indeed the way to reconstruct the a/s distinction: if there are strongly-CS content generalizations, they seem to detail the conceptual or analytic relations among contents, and vice versa. One is even tempted to give a demonstration: Suppose that 'B*#F# B*#G#' is such an S-generalization, then in every (nomologically) possible world, whenever someone comes to have an #F#-thought she tends to have a #G#-thought. Now suppose that 'F G' is not a conceptual/analytic truth. Then it is possible for someone to have an

⁴⁵ There is a certain sense in which the prototype theory of concepts can be seen to support the claim that people have outstandingly robust set of contingent beliefs surrounding particular concepts. But see Barsalou (1987) for strong evidence that prototypes are not robustly shared even intrapersonally, let alone interpersonally!

⁴⁶ It may be thought that a cluster theorist may accommodate this fact: what is required is not whether all the generalizations specified in the theory for a particular content* hold in the case of each subject, but rather that a substantial number of them do, while no particular generalization is necessary. This is also Stich's line on type-individuating syntactic objects. However, this can't be quite right. A cluster content functionalist must choose the S-generalizations in the cluster from among those that are already lawlike or, if you like, strongly-CS. But it usually so happens that each strongly-CS generalization about, say, bachelor-thoughts* is also necessary for having bachelor-thoughts*.

F-thought without ever being disposed to have a *G*-thought. But this seems to contradict the original assumption that ‘*B**#*F*# *B**#*G*#’ is such a strongly-CS generalization, supposing that ‘*B**#*F*# *B**#*G*#’ implements the intentional generalization, ‘*B*(*F*) *B*(*G*).’

Conversely, if there are conceptual/analytic relations among contents then they can be specified in terms of strongly-CS content generalizations. Intuitive quasi-proof: Suppose that ‘*F* *G*’ is a conceptual truth, then it follows that it is not possible for someone to have an *F*-thought without being disposed to have a *G*-thought. But this just means that ‘*B**#*F*# *B**#*G*#’ is a strongly-CS generalization, supposing, again, that ‘*B**#*F*# *B**#*G*#’ implements the intentional generalization, ‘*B*(*F*) *B*(*G*).’

If this is right, then, a localist functionalist might claim that the attempt to specify functional roles of tokens in terms of strongly-CS S-generalizations is the attempt to specify tokens’ analytic connections. Analyticity reduces to strong counterfactual support (or, perhaps, to lawlikeness, if you like).⁴⁷

As I said, I am not sure whether the proposal can be further clarified and sustained. My point, at this stage, is this: even if the proposal to distinguish between S-generalizations in terms of strong counterfactual support can be successfully elaborated and defended, and even if the distinction is shown to reconstruct the a/s distinction, and even if we assume, contra Fodor, that Quine’s criticism of the a/s distinction can be successfully met, there is the following obvious problem.

Remember that there must be *enough* interpersonally applicable and lawlike (= strongly-CS — in what follows I will use them interchangeably for the sake of brevity) generalizations to secure *unique* functional roles for each intuitively distinct content. Are there enough S-generalizations like that? If we assume, along with the localist functionalist, that many such proposed S-generalizations will mirror the conceptual/analytic connections among concepts *as traditionally/historically understood*, such as the intentional versions of the likes of (1) and (2), it is obvious that there aren’t enough of them to secure unique functional roles for almost any content.

⁴⁷ Devitt (1996) defends the notion of analyticity against Quineans. He claims that there is a principled distinction between those inferential links that are constitutive of the meanings of (mental) words and those that are not. He proposes that such links are all reference determining. But of course this alone can’t help distinguish those meaning-constitutive inferential links from those that are not, since, for instance, ‘tiger \nrightarrow dangerous’ is also reference determining in that whatever ‘tiger’ refers to ‘dangerous’ also refers to. He needs a modal operator to distinguish between, say, ‘bachelor \nrightarrow unmarried’ and ‘bachelor \nrightarrow neurotic’ (suppose, for the sake of the example, that psychologists discovered that all bachelors are neurotic, and this became a widely believed view). If I understand him correctly (in personal communication), his proposal is to appeal to something very much like what I tried to capture with my discussion here: (1) holds in the strongest metaphysical sense of necessity (short of logical one), whereas (6) does not. He then tries to cash out the necessity involved by appeal to the metaphysical structure of the world. See also his (1991b).

Interestingly enough, this point is acknowledged by some functional role semanticists. Block, for instance, writes, in a recent polemical article written against Fodor and Lepore's criticism, thus:

Fodor and Lepore seem to assume... that... the inferential role theorist has the option of appealing to analyticity as a way of discriminating the inferential liaisons that are in inferential roles from those that are out. But if we stick to traditional ideas about the extension of 'analytic', there aren't *enough* analyticities. Consider the putative analytic truths involving 'cat' — 'Cats are animals', 'Cats are living beings', 'Cats are grown up kittens', etc. The problem is that abstracting from the words 'cat', 'kitten', etc., appearing in these sentences, there is nothing here to distinguish 'cat' from 'dog'. Corresponding to 'Cats are grown up kittens', we have 'Dogs are grown up puppies'. Sure, 'nothing is both a cat and a dog' can be used, but so can 'nothing is both a dog and a cat'. Even if 'Cats are feline', and 'Dogs are canine' are analytic, this is of no help without other analytic truths that distinguish 'feline' and 'canine'... (Block 1993: 3–4)⁴⁸

Trying to secure a unique functional role for every token on the basis of S-generalizations that detail "conceptual" relations among them is like trying to *define* every concept or term. But even if it were possible to give a few necessary/analytic connections for some concepts or even for each concept, it would be overly optimistic, to put it mildly, to claim that each concept could be given both necessary and sufficient conditions for its application, especially after observing, as Fodor once put it, the failure of philosophy to define any term of any significance after two millennia. Similarly, there is no reason to expect that each token (with a distinct meaning) can be secured a unique functional role in terms of S-generalizations like (1) and (2).

It is, then, the pressure to secure unique functional roles that forces the local functionalist to a holistic position. To secure uniqueness, there must be enough generalizations. The scarcity of "analytic" generalizations forces the functionalist to look for other sorts of generalizations that pretty commonly hold among the folk but specify causal transitions between states whose representational contents, intuitively put, are only contingently connected. (4), (5) (and, even 3) may be examples to these kinds of S-generalizations. Hence the move to a ND-holist functionalism.

So let us take up ND-holism again. There are now three problems that such a functionalist must solve. One is that T now will contain S-generalizations that don't seem to be lawlike or strongly-CS. As I said, such generalizations as (4) and (5) seem to detail the causal profile of tokens whose representational contents are only contingently connected. As such they are like

⁴⁸ See also Block (1986:628–9); Cf. Loar (1982a:81ff.).

commonly held ordinary empirical beliefs. And the inferential relations among such beliefs don't seem to be a matter expressible in terms of nomological necessities. Put this worry temporarily aside. It may be that the functionalists' initial requirement that the generalizations that go into T be lawlike was just too strong. Perhaps mere counterfactual support is enough for them to go beyond being descriptions of mere statistical summaries of what typically causes what. If so, it was perhaps a mistake to demand that they be lawlike.

The second problem, however, is more serious. As I characterized ND-holism, what makes it non-destructive is that the S-generalizations are applicable interpersonally. But how do we determine how big their application domain actually is or ought to be. I said that their domain should be *sufficiently* large to cover all those with a reasonably common intentional psychology. Even if we ignore the potential circularity in specifying the domain in such intentional terms as 'common intentional psychology,' it is the problem of vagueness which seems to be so bothersome and intractable. How many people are enough for an S-generalization to count as meaning-constitutive? Surely there is no principled boundary to be drawn by just counting heads here. Furthermore, even if we come up with a set of S-generalizations whose scope is sufficiently large to secure unique roles, is there any guarantee that this will be the case for each concept* at any given time?

Relativization to particular times is also a serious problem here since there is no guarantee that the set of such generalizations will not exhibit variations over time, both in terms of the number of heads they cover and in terms of admitting "new" and dropping "old" generalizations in the set. In fact, given that we dropped the requirement of lawlikeness as characterized above, there is every reason to believe that the S-generalizations in T, even though may have a sufficiently large population as their domain, will exhibit variations over time. This is intuitively clear if we think of the situation before, suppose, psychologists' discovery that all bachelors are neurotic, and after. Consider also the period during which the belief* that #all bachelors are neurotic# becomes popular, getting out of psychologists' labs and spreading to an entire population. Consider also skeptics, disbelievers, etc.

This brings us to the third problem, which is that even with all such pluralist S-generalizations (with sufficiently large scope for a particular time period) included in T, there is *still* no guarantee that a unique functional role will be secured for each distinct concept. It is obvious that the main pressure here is to secure uniqueness: as far as this is the sole non-negotiable constraint on individuation of functional roles, there is no guarantee that the S-generalizations covering a token will not become increasingly narrower in scope of application, and at the end, restricted perhaps only to single individuals. In fact, there is every reason to believe that they will.

This worry is exactly what lies at the source of Fodor's pessimism that such content functionalism inevitably leads to a destructive sort of holism, since there doesn't seem to be a robust and stable way of individuating functional roles *interpersonally* and in such a way that would also guarantee their projectibility (their appropriateness for genuine lawlike generalizations). With the move to accepting S-generalizations into T that are not lawlike and the desire to secure uniqueness, the functionalist moves into a position into which a serious vagueness is essentially built. This is already disturbing holistic terrain, since it is the first and apparently inevitable step taken towards specifying functional roles in terms of *all the available* S-generalizations including the singularist ones. So the slippery slope to a destructive holism is grounded on the constraint for uniqueness.⁴⁹

The trouble, then, arises in trying to simultaneously satisfy the three constraints on the S-generalizations: lawlikeness (or, being strongly-CS), uniqueness and interpersonal applicability. The solution, under the pressure of securing unique functional roles, is to appeal increasingly more to those S-generalizations whose scope is increasingly narrower and at the end restricted to single individuals. As we have seen, this also results in forcing the functionalist to sacrifice the constraint that generalizations used in the fixation of a functional role be strongly-CS or lawlike. This is, then, how you get a destructive sort of holism from localist (as well as ND-holist) functionalism. Localist functionalism all by itself doesn't seem to suffice to specify unique functional roles for content that are robustly applicable across people. But if localist functionalism is abandoned, lawlikeness is abandoned with the move to ND-holism, and then the pluralist scope of S-generalizations is given up, which leads exactly to holism of the sort Stich was accusing the content theorist of.

Now, as I said in the beginning, this is not quite the way Fodor, or for that matter any one else, characterizes the discontent with functionalism. Fodor's early attacks against holism assumed the truth of the second premise of the Ur-argument. But instead of directly arguing for it Fodor left the issue as if it were already established by Quine's attack on a/s distinction. As I tried to show above, defending (ii) on Quinean grounds is not available to Fodor, nor to a (holist) functionalist. Its defense must be conducted on non-intentional grounds. And this is what I've tried to do above.

Now with this characterization of content functionalism and of its troubles at hand, let us go back to my claim that the grounds for rejecting content functionalism are the same with rejecting functionalism in individuating Mentalese tokens across systems, hence rejecting Stich's STM.

⁴⁹ Indeed Block (1991, 1993) seems to take this line more or less explicitly.

If what I have said so far is correct, it is relatively easy to see how exactly parallel problems plague NCA of typing Mentalese tokens (=brain states with particular syntactic objects mapped onto them). Indeed the two enterprises are in fact identical as far as S-generalizations are concerned. The reason is that according to the content functionalist the (narrow) semantic content of a token is *metaphysically constituted* by the functional role it plays, and the job of specifying this functional role is exactly the job of type-individuating the token on the basis of its functional role as uniquely specified by S-generalizations. In fact, as I hinted above in my first clarificatory point about inferential relations, the content-functionalist needs a vocabulary to state the S-generalizations. I said that this vocabulary cannot be intentional for fear of violating the demands of naturalism. So the vocabulary in stating S-generalizations ought to pick up symbols as they are non-intentionally characterized. The device with '#’s and '*’s I used above in giving examples for S-generalizations in a way conceals the urgency of the problem by implicitly using the semantics of English. Take, for instance, the S-generalization (1) above:

- (1) For all *S*, if *S* comes to believe* that #*x* is a bachelor#, then, *ceteris paribus*, *S* will come to believe* that #*x* is unmarried#;

Putting aside the problem of computationally specifying what believing* is supposed to be, the functionalist needs to answer the question of how '#*x* is a bachelor#' and '#*x* is unmarried#' refer to what they are supposed to refer to. Their referents are supposed to be Mentalese symbol types that are individuated across systems on the basis of their non-intentional, and presumably, non-(quasi)-physical, properties that obey a certain set of causal/computational regularities. Such a vocabulary is not generally available to a functionalist prior to having already established these regularities on the basis of which she proposes to type-individuate Mentalese tokens, i.e. on the basis of their narrow functional profile across systems — see the next section.

The trouble with content functionalism is the trouble of getting a robust and stable individuation criterion for functional roles that are interpersonally applicable. If this is right, it is obvious that exactly the same worry must attach to the project of individuating syntactic Mentalese tokens across systems on the basis of their functional roles. Indeed, the projects are identical: while a content functionalist wants to assign a content, say, [water is wet], to a Mentalese token on the basis of its interpersonally applicable functional role, an STM-like functionalist wants to assign to it the syntactic type #water is wet# on exactly the same basis.

I think it must now be transparent why the job of assigning a content to a token on the basis of its functional role is exactly the same as the job of type-individuating it on the basis of its functional role. Indeed, the content is supposed to *metaphysically reduce* to the functional role

(non-intentionally characterized). This is after all what naturalism is supposed to be on a functionalist research program. I conclude, therefore, that to the extent to which narrow content functionalism has the problems I claimed it does to that extent the NCA-cum-STM has exactly the same problems. Stich is promoting a scientific framework for psychology that lacks the resources to type-individuate the relevant mental/brain states in a robust and stable manner in a way that respects interpersonal applicability of its theories.

6.2 Connections to Stimuli and Behavior

Before ending this section, we need to address one more issue. Although I characterized NCA as involving three kinds of generalizations detailing causal relations of a symbol token to stimuli (I-T laws), other symbol tokens (T-T laws), and a proprietary set of behaviors like basic motor-gestures (T-O laws), in my discussion of content functionalism I addressed only the S-generalizations that specify relations among tokens. It may be thought that introducing generalizations that specify nomological relations of tokens to stimuli and motor-gestures would help in securing unique roles. This is a natural reaction, but it won't help the defender of NCA.⁵⁰

Take, first, the causal generalizations that are supposed to lawfully connect a set of stimuli to, say, #Clinton is not faring well#, or any similarly specific sentence. Whatever the laws of psychophysics may tell us with respect to a very restricted range of psychophysically available properties, they will certainly be silent for the vast majority of symbol types figuring in full-blown propositional attitudes*. The problem partly stems from stimuli being *proximal*. There are certainly no scientifically well-delineated sets of proximal stimuli nomically correlated with the Mentalese tokens implementing propositional attitudes. This is to say that no such set could constitute a natural kind which would lawfully correlate with the computational implementers of propositional attitudes. The other part of the problem is the holism involved in belief* fixation. Which proximal stimuli will cause which symbol(s) to be tokened in the B-box is determined by what other symbols actually happen to be there and by the overall internal organization of the B-box (simplicity, conservatism, etc.).

⁵⁰ I should emphasize that similar issues can be raised with narrow content functionalists. Many narrow functionalists acknowledge that connections to proximal inputs and outputs play an indispensable role in individuating mental state tokens. However, most of the time, they ignore them right after they emphasize their importance and go on to discuss S-generalizations only. They just don't focus on the inputs and outputs. For a forceful criticism of this bad habit and its consequences, see Devitt (1990, 1991a). In their discussion, Stich (in his 1983), Fodor, Block are no exception to this habit. As I said before, Stich (1991:245) distinguishes between "fat syntax" (the type-identity of a symbol as determined by all its three kinds of relations) and "skinny syntax" (the type-identity of a symbol as determined only by the S-generalizations covering it) and opts for the former as the true characterization of his NCA-cum-STM.

The history of behaviorism also provides an overwhelming inductive evidence that there are no such laws to be stated. No one has ever succeeded actually stating a single such law! Similarly for the supposed generalizations that would lawfully connect basic motor-gestures to particular symbol types in the B- and D-boxes. To be sure, behaviorists were after lawful stimuli/behavior connections, which is different. But the moral must be the same, since their failure primarily stemmed from an inability to find projectible predicates to apply to all and only those proximal stimuli *under physical descriptions* that lawfully govern a given piece of behavior. They assumed that such stimuli directly and lawfully control the relevant piece of behavior: they wanted to bypass mediating internal states. They failed primarily because of the holism problem again. Nothing seems to change, however, if we assume that it is particular propositional attitudes*, rather than behavior, that are directly under the lawful control of proximal stimuli: the routes from stimuli are equally holistic in each case. Here is how Fodor makes much the same point:

I wanted to say that P(INF) [the “name for the disjunction of all the proximal stimuli which can cause “horse” to be tokened” in one’s B-box] is an open disjunction and that properties that are expressed by open disjunctions don’t enter into laws. (In fact, given that tokenings of “horse” are often theory mediated, P(INF) probably includes every proximal stimulus since, as I remarked in *TOC* [1990: 108–10]..., the merest ripple in horse infested waters can produce proximal stimuli which cause “horse” tokenings in the mind of a properly informed observer.) (1991:256, Reply to Antony and Levine)

Perhaps I am laboring this point needlessly. It should be clear that there are no lawlike generalizations to be stated with respect to proximal inputs/outputs for the full range of particular symbol types deployed in central cognitive processing as direct objects of propositional attitudes*.⁵¹ And even if there may happen to be some, they will be so few and fragile that they

⁵¹ It is very curious that more or less the same criticism is given by Stich himself against content functionalists’ alleged claim that there are such generalizations: “[t]here is generally no characteristic environmental stimulus which typically causes a belief. There is no bit of sensory stimulation which typically causes, say, the belief that the economy is in bad shape, or the belief that Mozart was a freemason... Nor do most beliefs have typical behavioral effects. My belief that Ouagadougou is the capital of Upper Volta does not cause me to do much of anything” (1983:24). Later on, he argues (1983:180–1), on familiar grounds, that there can be no principled distinction between beliefs whose content is “observational” and those whose content is “theoretical.” So, according to Stich, even for allegedly “observational” beliefs there doesn’t seem to be any particular set of stimuli nomologically connected to them. I wonder why and how Stich could think that the parallel case of beliefs* with particular symbol types as their objects is immune to a parallel criticism!

will be of very little help in type-individuating all the symbol types we may need in psychological explanations.

It is clear, then, that the heaviest burden for the individuation of symbol tokens must be carried by the S-generalizations that hold among particular symbol types. But we have already seen that individuation on the basis of S-generalizations seems hopeless without succumbing into a destructive sort of holism.

I am using the holistic nature of *belief* fixation as an argument against Stich, and in a way switching implicitly back and forth between beliefs and B-states. Now, of course, B-states mapped to particular syntactic objects are *not* beliefs with particular contents in any obvious sense. So, Stich might say, this kind of criticism can at best be leveled against (narrow) content functionalism and not against STM. But, really, why not? Granted that B-states are not beliefs, what reasons do we have to think that fixation of B-states initiated by sensory stimulation is not holistic? Surely, Stich has not provided us with any such reason. In fact, given that the cases are exactly parallel, and that B-states are the states that respond to stimuli and interact with other central states like D-states to produce behavior, *prima facie* we have every reason to believe that the links between B-states, and stimuli or behavior are equally tenuous.

It is important to emphasize that we are talking about B-states that interact with other similar central states to produce behavior, and the I-T generalizations are supposed to cover *these* states mapped to particular syntactic objects. The question I am raising is whether there is any *such* generalization nomologically connecting proximal stimuli to a B-state with a particular syntactic object in virtue of which it presumably interacts the way it does with other states having other syntactic objects assigned to them to produce behavior. It is important to be clear about this. For, presumably, there are a lot of nomological connections between types of stimuli or motor gestures and certain neurological states that may not be B- or D-states in an obvious sense. For instance, psychophysics is in the business of discovering many connections between stimuli and some states of the nervous system. But I don't think that Stich would want to say that the immediate output state of, say, a sensory transducer is a B-state (or, whatever it's that an STM "*cognitive*" theory would want to posit as theoretically central brain state types). Surely the burden of the proof is on Stich.

Let me indicate two more problems for the kind of type individuation of mental sentences Stich envisages for STM. As remarked, Stich proposes a cluster view of identifying neural states as particular sentence tokens: "to count as a token of a sentence type, a neurological state must satisfy some substantial number of the cluster of generalizations included in a theory, without specifying any particular generalizations that must be satisfied, nor exactly how many must be satisfied" (1983, p.152). He admits that this introduces vagueness into the iden-

tivity conditions of mental sentences. However, the problem this may cause is more than just introducing vagueness. It risks downright misclassification. Consider again S-generalizations, since, in a certain sense, they are expected to do the heaviest work in the individuation of sentences on the STM framework. The problem is that there may be two sentence tokens satisfying almost the same generalizations but nevertheless differing in type because they satisfy a few different “essential” generalizations. Consider the token belief* that #...gay...# and the token belief* that #... lesbian ...#, it is likely that they have very similar causal roles. What may be distinguishing them are just a few (counterfactual supporting) S-generalizations such as ‘B*(#gay#) B*(#male homosexual#)’ and ‘B*(#lesbian#) B*(#female homosexual#)’. What reasons can Stich give us that such cases are not seriously troublesome or do not really arise? I can see none.

Another problem is one that Stich again himself raised against content functionalism. On Stich’s own admission, given two subjects with the same B- and D-states, the potential as well as actual causal patterns (concerning especially the ones captured by L-generalizations) that their B- and D-states will exhibit are very likely to differ. This is the problem parallel to the one that the content functionalism faces: the kinds and the degree of complexity of inferences that people can draw vary greatly from person to person. If any attempt to incorporate these different causal patterns into a functionalist theory in a principled way will be, as Stich says, “ad hoc and implausible”, how can Stich think that an STM-theorist’s parallel attempts will not similarly be *ad hoc* and implausible? Notice that here insisting that B-states are not beliefs cannot even begin to help: the explanation of a certain kind of mental activity on the basis of purely syntactic transformations of some complex abstract objects mapped onto B- and D-states is exactly what STM theories are supposed to be good at. Stich, of course, would like to say that for such cp-dissimilar subjects, an STM-theorist will specify different L-generalizations. But why can’t the content psychologist do the same thing? (See § 5.2)

This concludes my second argument against STM. I hope to have shown that the NCA of typing brain sentence tokens is bankrupt beyond repair. It cannot type individuate particular B-states without committing itself to a destructive sort of holism.

I have already hinted at the argument for my third claim above. Now let me take it up explicitly: no STM-theory can get off the ground without using intentional idioms.

7. Why a Purely Syntactic Psychology Cannot Get off the Ground

Throughout Stich’s 1983 book, there are various passages in which Stich seems to argue that an STM-theorist had better refrain altogether from using intentional notions even in the theory construction stage. Here is a typical one: “cognitive psychologists can and do develop the theory

of mental processes without attending to the semanticity of formulae in the mental code” (1983: 193). In fact, Stich’s discussion of what he calls the Weak RTM is an attempt to show that assuming that the formulae have semantic content is frivolous at any stage of theory development.

Many people seem to think that functionalism in scientific psychology can be, to a significant extent, carried out without ever raising any semantical worries. In this section, I will argue that this in fact can’t be done. In particular, I will show that construction of an STM-style theory cannot be carried out without using intentional notions. This problem is one that seems to belong to “the context of discovery”, but nevertheless, I think, it will be instructive to see why an STM-theorist is committed to using intentional notions in at least theory construction stage. I already detailed the reasons why STM is seriously problematic otherwise.

In fact, it is for a very simple reason that within a strictly STM paradigm theory construction cannot get off the ground without using any intentional idioms. STM is a purely functional theory. As such, all the theoretical predicates that denote functionally defined particular brain state types depend for their reference on the entire theory being *in situ*. In other words, within the STM paradigm, the only legitimate way to refer to the nodes of the causal network of brain states is by way of theoretical terms whose applicability entirely depends on the theorist’s having almost the whole functional theory first. That is his point when Stich insists that the type identity of a sentence token (a brain state token) entirely depends which and how many generalizations cover it.

It is only against the background of a systematic mapping of state types to sentence types that any given state token counts as a token of a particular sentence type... No one neurological state can count as a token of a sentence type unless many neurological states count as tokens of many different sentence types. But this holism... is quite distinct from the holism imposed on the folk psychological notion of belief by the embedded appeal to ideological similarity. For the status of a state as a token of a sentence does not depend on what other cognitive states a subject currently happens to be in. It depends only on the causal interactions that the state would exhibit with stimuli, with behavior, and with other states. (1983:153)

But there is no way to start theory construction without having an initial and independent way of referring to the nodes of the causal network of the brain states about which nothing is known in the initial stages. In other words, when there is no theory yet, the prospective “theoretical terms” can’t refer. In the initial stage of theory construction, the theorist has no idea whatsoever what, say, ‘the B-state mapped to Fa’ refers to. This presents a dilemma. On the one hand the STM-theorist wants to theorize about the functional organization of particular brain

state types. For this she must have an independent way of referring to them, independent of a more or less completed theory. On the other hand, as far as she refrains from using an intentional scheme, she can't even guess what she is talking about when she uses terms like 'the B-state mapped to Fa'. That is because the theorist has no independent way of identifying the nodes of the network of brain state types. This network is completely unknown.

The problem stems from the STM paradigm itself. Notice that if there were an independent way of picking up the nodes (brain state types) in the causal network that does not presuppose a more or less complete specification of what those nodes are connected to what other nodes and how (i.e. their potential functional roles determined by the generalizations of the theory), then we would use this scheme in our way to saying what generalizations there are, i.e. in our theory construction. This is exactly what Brian Loar does (1982a) in presenting his semi-broad content functionalism: he uses propositions to pick up those brain states and state whatever generalizations there are that need to be stated. Once he gets the generalizations, he gets rid of propositions in favor of syntactic objects. Then, of course, he is in a position to specify, theoretically at least, all the functional roles there are without using any semantic terms. Once he does that the result is almost an STM theory very much like what Stich envisions.

So it should be obvious that the way out of this dilemma can only be semantic, not syntactic. The upshot is that a purely "syntactic" (psycho)functionalism in scientific psychological theory construction à la Stich cannot be carried out without assuming the truth of content (semantic) functionalism (à la Loar). They stand or fall together, which is not to say that narrow content functionalism has got to be true (see above).

If what I have said so far is right, the lesson to be drawn is that syntactic functionalism is not an option in psychological *theory construction* somehow at the discretion of a psychologist. When we reflect upon the historical rise of functionalism in the philosophy of mind, that this is so should be obvious. Functionalism was developed as a response to the inherent difficulties in behaviorism and (type-type) identity theories. It was conceived as a metaphysical theory saying what mental states are. Functionalism identified mental states with functional states. But that was not enough. Functionalism had to be able to provide the identity conditions of mental state types. This required providing identity conditions for functional roles. Functionalists had to be able to say what functional roles uniquely define what types of mental states. But this required having a theory first. Some versions of functionalism took this theory to be folk psychology made explicit with all the intentional/mental terms employed as theoretical terms. Then, Ramsefying this theory was the major step in explicitly getting the identity conditions for functional roles. Similarly with psychofunctionalism: the theory to be Ramsefied was conceived to be a theory to be developed by scientific psychology. The underlying idea was the same. Once such a theory was at our disposal with all its intentional/mental terms employed as theoretical

terms, we could explicitly get the identity conditions for the functional roles by Ramsefying it. In all this, the construction of the theory to be Ramsefied was conceived along with using all the intentional vocabulary available to the theorists. And that was OK, because functionalism was competing against dualism, eliminativism and reductionism (type identity theory). That is the reason why functionalism at its core is essentially an intentional realist theory. But Stich's STM tries to reverse the situation, it wants to develop functional theories without ever using intentional terms; in this, however, Stich is putting the cart before the horse. As we have seen, this turns out to be practically impossible, because the remaining vocabulary to be used in theory construction cannot do the required job. In a sense, in fact, Stich cuts the branch he is sitting on.

Admittedly, my point in this section is one that belongs to the context of discovery. It might be claimed that as such it is not that important: what matters is whether the ultimate STM-style theory, when completed, is committed to any intentional scheme. The STM theorist might use any tools (intentional or otherwise) that would help in getting the theory, i.e. in the context of discovery. But once the theory is completed and successful, it should not matter how it was gotten in the first place. For instance, as long as it belongs to a discovery stage, an STM-theorist might use a procedure like Loar's. It is the form of the ultimate completed theory that counts.

Well, I have two points to make on this. First, given Stich's criticism of content-based psychologies, it should be obvious that the brain states initially typed according to an intentional scheme will exhibit all the vagueness, context-sensitivity, and parochialism that Stich claims will pertain to a semantic taxonomy. So he can't avail himself to the SA of typing even in the context of discovery. Second, it is simply absurd to assume that a taxonomic scheme will be semantic free if at the end it is essentially obtained by a SA and then gotten rid of à la Loar. The ultimate theory, if really successful, is nothing but a (partial) scheme for a naturalized semantics (e.g., in the tradition of two-factor semantic theories).⁵²

8. If Cognition Is Computational, How Can Psychological Laws be Intentional?

[[The material in this section needs to be elaborated in a more comprehensive and argumentative way especially in the context of a more general discussion of mental causation — in progress...]]

This is what Fodor called the "Eponymous Question" in his (1996). This question has in fact been around, constantly popping up here and there and haunting people working in the field, for more than ten years, mostly thanks to Stich, as we have seen. [See, among many others,

⁵² On this last point, see also Higginbotham (1988).

Stich (1983, 1991), Field (1978), Schiffer (1987), Fodor (1980, 1989, 1996), Devitt (1991a) who take issue with the EQ one way or other.]

This question is also related to some puzzles computationalism has created *vis-à-vis* mental causation. According to the computational picture of mind (CRT, LOTH), mental processes are defined over mental symbols physically realized in the brain. But computationalism says that for these mental processes to qualify as computational, it is the non-semantic, in particular syntactic, properties of symbols that the processes must be causally sensitive to. In fact, given a physicalist framework, it is not even clear what it would be like for mental processes to be causally sensitive to the semantic properties of symbols, which are relational, i.e., hold between the symbol (or the organism) and environment. Given the locality of causation, all thought processes can be causally sensitive to seem to be the syntactic (at any rate, non-semantic) properties of symbols that are implemented neurally. If so, even though mental symbols are causally efficacious in reasoning and causation of behavior, it is in virtue of having certain syntactic properties, but not in virtue of having semantic properties, that they are so. Thus, as far as the science of psychology is in the business of causal explanation, the relevant properties of mental states in virtue of which they are covered by causal psychological laws are all non-semantic, or so it seems. This is another way of seeing Stich's motivation in arguing against content-based psychologies and promoting his STM over them. As we have seen, Stich calls the Narrow Functional Account of typing symbol tokens "syntactic" typing, presumably meaning just non-semantic and non-physical.⁵³ And this sort of typing, on his view, is what the STM (or CTM for that matter) is committed to. He then claims that STM/CTM is all a scientific psychology needs; hence, contra Fodor, no need to appeal to semantic/intentional properties of syntactically structured brain symbols in stating the laws of psychology. He accuses Fodor to have it both ways.⁵⁴ We are now in a position to see how it is possible to have it both ways, i.e., to see what the answer is to the Eponymous Question.

Let us suppose that computational psychology is correct. Any scientific computational psychology needs to postulate states in terms of which it can explain (and predict) behavior (construed broadly —bodily, verbal, mental behavior). This seems to call for covering laws or generalizations that subsume those states under an appropriate description. This means that these states, under the relevant description, are projectible, i.e. natural kinds from the perspective of the theory. As such they must have identity conditions. Computational psychology characterizes these states as symbol tokens realized in the heads of cognitive organisms. *Qua*

⁵³ Using 'syntactic' in this sense is at best misleading. See my (1997) for an extensive discussion of the notion of syntax required for LOTH.

⁵⁴ Stich (1983). See also his (1991). Devitt (1991a) joins Stich in accusing Fodor of trying to have it both ways but only with respect to processes governing thoughts without I/Os.

symbols they have both syntactic and semantic properties. OK then, how are we to type them to suit the psychological laws covering their tokens? We have seen that they cannot be typed, in the required sense, by their narrow functional properties: NFA is hopeless. The Physical Account (PA) of typing them is hopeless too. Stich and almost everybody in the field agrees. PA seems to commit one to a very strong version of type-type identity theory for propositional attitudes with specific content (like the belief *that snow is white*) cast across people. In this form, PA has almost no defenders. Our only other option, then, the Semantic Account, is in fact mandatory if psychological processes are to be computational. In other words, if Stich's original question, i.e. the question of what it is for two symbol tokens of Mentalese in different heads to be of the same type, has an answer, it must be some version of SA.⁵⁵ It must be on the basis of their semantic properties we type symbol tokens across systems.

I therefore conclude that computational psychology (CTM, for that matter) itself is essentially committed to semantic type individuation of symbol tokens across systems. And it is across systems that a scientific psychology casts its laws. Hence, the necessity for intentional psychology whose laws advert to the semantic properties of representations. If mental representations can be typed interpersonally *only* on the basis of their semantic properties, CTM cannot be an alternative to replace intentional psychology. Hence the answer to Fodor's Eponymous Question.⁵⁶

⁵⁵ It is of course possible that Stich's question doesn't have an answer. I surely haven't argued here independently for the truth of SA. In other words, if Stich is right about the fate of SA, and if I am right about the fate of PA and NFA, then scientific cognitive psychology as we know it today is impossible. I can't take this option seriously, in particular I can't take seriously a priori arguments against the cogency of the foundations of what appears to be an enormously successful and fruitful scientific approach to cognition. Cognitive psychology seems to be into intentional talk up to its neck. I take it that there is an enormous *prima facie* evidence for the truth of the intentional assumptions of present day cognitive psychology. I take this to be the best argument for SA albeit a non-demonstrative one. I left Stich's positive arguments against SA aside in the beginning of the paper. What needs to be done, of course, is to address Stich's criticisms in order to begin to give an independent argument for SA.

⁵⁶ There are, to be sure, problems with any version of SA, as is well known. Suppose that SA is broad as in Fodor. Then we have problems with Frege cases as well as Twin-Earth cases. A narrow SA would be equally problematic, as we have seen, if it relies on narrow functional roles of vehicles as their narrow semantic content. On the other hand, a Fodor-style notion of narrow content as mapping from context to broad content can perhaps handle Twin-Earth cases at best, but not Frege cases (see my 1996). But being problematic is one thing, being wrong is another: I think, a SA that works can after all be salvaged in the face of apparent difficulties, see my (in prep.).

REFERENCES

- Aydede, Murat (1996), "On the Type/Token Relation of Mental Representations," to appear in *Facta Philosophica*.
- Aydede, Murat (1997), "Language of Thought: The Connectionist Contribution," *Minds and Machines*, Vol.7, No.1, pp.1–45.
- Aydede, Murat (in prep.), *Computation, Representation and the Mind*, The University of Chicago.
- Barsalou, L. W., (1987), "The Instability of Graded Structure: Implications for the Nature of Concepts" in *Concepts and Conceptual Development*, U. Neisser (ed.), Cambridge, UK: Cambridge University Press.
- Block, Ned, (1978), "Troubles with Functionalism" in *Readings in Philosophy of Psychology*, N. Block (Ed.), Vol.1, Harvard University Press, 1980.
- Block, Ned (Ed.), (1980a), *Readings in Philosophy of Psychology*, Vols.1 & 2, Harvard University Press.
- Block, Ned, (1986), "Advertisement for a Semantics for Psychology" in *Studies in the Philosophy of Mind: Midwest Studies in Philosophy*, Vol.10; French, P., T. Euhling and H. Wettstein (Eds.), University of Minnesota Press, Minneapolis, 1986.
- Block, Ned, (1991), "What Narrow Content Is Not" in *Meaning in Mind: Fodor and his Critics*, Loewer and Rey (Eds.), Basil and Blackwell, 1991.
- Block, Ned, (1993), "Holism, Hyper-analyticity and Hyper-compositionality," *Mind and Language*, Vol.8, No.1, pp.1–26.
- Crane, Tim, (1990), "The Language of Thought: No Syntax Without Semantics," *Mind and Language*, Vol.5, No.3, pp.187-212.
- Devitt, Michael, (1990), "A Narrow Representational Theory of the Mind," in *Mind and Cognition*, W.G. Lycan (Ed.), Basil Blackwell, 1990.
- Devitt, Michael, (1991a), "Why Fodor Can't Have It Both Ways" in *Meaning in Mind: Fodor and his Critics*, B.Loewer and G. Rey (Eds.), Basil Blackwell, 1991.
- Devitt, Michael, (1991b), "What Did Quine Show Us about Meaning Holism?", Draft, University of Maryland, College Park.
- Devitt, Michael, (1996), *Coming to Our Senses: A Program for Semantic Localism*, Cambridge University Press.
- Devitt, Michael and Kim Sterelny, (1987), *Language and Reality*, The MIT Press, Cambridge, Massachusetts.
- Field, H. (1978), "Mental Representation," *Erkenntnis* 13, 1:9–61.
- Fodor, Jerry A., (1975), *The Language of Thought*, Harvard University Press: Cambridge, Massachusetts.

- Fodor, Jerry A., (1980a), "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology" in *RePresentations: Philosophical Essays on the Foundations of Cognitive Science*, J. Fodor, The MIT Press: Cambridge, Massachusetts, 1981. (Originally appeared in *Behavioral and Brain Sciences* 3, 1, 1980.)
- Fodor, Jerry A., (1980b), "Methodological Solipsism: Replies to Commentators", *Behavioral and Brain Sciences* 3, pp.99-109.
- Fodor, Jerry A., (1983a), "Reply to Brian Loar's 'Must Beliefs Be Sentences?'" in *Proceedings of the Philosophy of Science Association for 1982*, Asquith, P. and T. Nickles (Eds.), East Lansing, Michigan, 1983.
- Fodor, Jerry A., (1983b), *The Modularity of Mind*, The MIT Press: Cambridge, Massachusetts, and London, England, 1983.
- Fodor, Jerry A., (1985), "Fodor's Guide to Mental Representation: The Intelligent Auntie's Vade-Mecum", *Mind* 94, 1985, pp.76-100. (Also in TCOE.)
- Fodor, Jerry A., (1987), *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, The MIT Press: Cambridge, Massachusetts, 1987.
- Fodor, Jerry A., (1989), "Substitution Arguments and the Individuation of Belief" in *A Theory of Content and Other Essays*, J. Fodor, The MIT Press, 1990. (Originally appeared in *Method, Reason and Language*, G. Boolos (Ed.), The Cambridge University Press, 1989.)
- Fodor, Jerry A., (1990), *A Theory of Content and Other Essays*, The MIT Press.
- Fodor, Jerry A., (1991), "Replies" (Ch.15) in *Meaning in Mind: Fodor and his Critics*, B. Loewer and G. Rey (Eds.), Basil Blackwell, 1991.
- Fodor, Jerry A., (1994), *The Elm and the Expert*, The MIT Press.
- Fodor, Jerry A. and Ernest Lepore, (1991), "Why Meaning (Probably) Isn't Conceptual Role?", *Mind and Language*, Vol.6, No.4, pp.328-43.
- Fodor, Jerry A. and Ernest Lepore, (1992), *Holism: A Shopper's Guide*, Blackwell, 1992.
- Higginbotham, J., (1988), "Is Semantics Necessary?", *Proceedings of the Aristotelian Society* 88, pp.129-41.
- Kripke, S.A., (1980), *Naming and Necessity*, Harvard University Press.
- Lewis, David (1973). *Counterfactuals*, Oxford, UK: Blackwell.
- Loar, Brian F., (1982a), *Mind and Meaning*, Cambridge University Press, 1982.
- Loar, Brian F., (1982b), "Must Beliefs Be Sentences?" in *Proceedings of the Philosophy of Science Association for 1982*, Asquith, P. and T. Nickles (Eds.), East Lansing, Michigan, 1983.
- Loewer, Barry and Georges Rey, (Eds.), (1990), *Meaning in Mind: Fodor and his Critics*, Oxford: Basil Blackwell.
- Lycan, William G. (Ed.), (1990), *Mind and Cognition: A Reader*, Basil Blackwell.

- Putnam, Hilary, (1975), "The Meaning of "Meaning"" in Gunderson, K. (Ed.), *Minnesota Studies in the Philosophy of Science*, Minneapolis, University of Minnesota Press, 7:131-193.
- Ramsey, William, Stephen Stich and Joseph Garon, (1990), "Connectionism, Eliminativism and the Future of Folk Psychology," *Philosophical Perspectives: Action Theory and Philosophy of Mind*, 4.
- Schiffer, Stephen, (1987), *Remnants of Meaning*, The MIT Press.
- Stalnaker, Robert L., (1968). "A Theory of Conditionals," *Studies in Logical Theory*, *American Philosophical Quarterly*, Monograph Series, No.2, Oxford, UK: Blackwell, pp.98-112.
- Stich, Stephen P., (1978), "Autonomous Psychology and the Belief-Desire Thesis" in *Mind and Cognition*, W.G. Lycan (Ed.), Basil Blackwell, 1990. (Originally appeared in *The Monist* 61, pp.573-591, 1978.)
- Stich, Stephen P., (1983), *From Folk Psychology to Cognitive Science: The Case Against Belief*, The MIT Press: Cambridge, Massachusetts; London, England.
- Stich, Stephen P., (1988), "Connectionism, Realism, realism," *Behavior and Brain Sciences*, 11:3. [Comment on Dennett's "Precis of *The Intentional Stance*"]
- Stich, Stephen P., (1991), "Narrow Content Meets Fat Syntax" in *Meaning in Mind: Fodor and his Critics*, B.Loewer and G. Rey (Eds.), Basil Blackwell, 1991.