



CHICAGO JOURNALS



---

What's Wrong with the Syntactic Theory of Mind

Author(s): M. Frances Egan

Source: *Philosophy of Science*, Vol. 56, No. 4 (Dec., 1989), pp. 664-674

Published by: [The University of Chicago Press](#) on behalf of the [Philosophy of Science Association](#)

Stable URL: <http://www.jstor.org/stable/187786>

Accessed: 17/02/2011 14:42

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=ucpress>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).



The University of Chicago Press and Philosophy of Science Association are collaborating with JSTOR to digitize, preserve and extend access to *Philosophy of Science*.

<http://www.jstor.org>

**DISCUSSION:**  
**WHAT'S WRONG WITH THE SYNTACTIC THEORY  
OF MIND\***

M. FRANCES EGAN†

*Department of Philosophy  
Rutgers University*

Stephen Stich has argued that psychological theories that instantiate his Syntactic Theory of Mind are to be preferred to content-based or representationalist theories, because the former can capture and explain a wider range of generalizations about cognitive processes than the latter. Stich's claims about the relative merits of the Syntactic Theory of Mind are unfounded. Not only is it false that syntactic theories can capture psychological generalizations that content-based theories cannot, but a large class of behavioral regularities, readily explained by content-based theories, appear to be beyond their explanatory reach.

**1. The Syntactic Theory.** In his book *From Folk Psychology to Cognitive Science* (1983) Stephen Stich develops an alternative to the representational theory of mind defended by Fodor and Pylyshyn. The Syntactic Theory of Mind (hereafter STM) construes cognitive states as uninterpreted syntactic objects:

[C]ognitive 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 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 causal relations among cognitive states mirror formal relations among syntactic objects. If this is right, then it will be natural to view cognitive state tokens as tokens of abstract syntactic objects. (Stich 1983, p. 149)

According to Stich, a cognitive theory which instantiates the STM is to be articulated in three parts: (i) a specification of a finite set of primitive syntactic objects and a set of formation rules which can generate (infinitely many) complex formulae; (ii) the hypothesis that for each model

\*Received August 1987.

†I am grateful to Robert Matthews, Jerry Fodor, Ausonio Marras, and Ken Warmbrod for helpful comments on an earlier version of this paper.

*Philosophy of Science*, 56 (1989) pp. 664–674.  
Copyright © 1989 by the Philosophy of Science Association.

of the theory there exists a set of state types (presumably neurological state types) whose tokens are causally efficacious in the production of behavior, and further, that there is a mapping from these state types to syntactic objects in the class specified in (i); and (iii) a specification of the theory's lawlike generalizations. Such theories may plausibly be construed as constitutive of an *abstract neurology*, inasmuch as equivalence classes of neurological states are mapped into a space of syntactic objects; the equivalence classes are defined by the causal role of the states in the production of behavior.

The generalizations of STM theories, which characterize the causal relations among neurological states, are said to be "specified indirectly via the formal relations among the syntactic objects to which the neurological state types are mapped" (p. 151). Stich gives the following as a typical example of such a generalization:

For all subjects S, and all wffs A and B, if S has a B-state mapped to  $A \rightarrow B$  and if S comes to have a B-state mapped to A, then S will come to have a B-state mapped to B.<sup>1</sup> (p. 155)

STM theories, Stich argues, are to be preferred to theories whose generalizations advert to the *content* of mental states (that is, representationalist, or *RTM* theories), because they can (i) "do justice to" all of a content-based theory's generalizations, and (ii) capture *additional* generalizations which are beyond the reach of content-based theories. In this paper I shall challenge both claims. I shall begin with the latter claim, arguing that generalizations beyond the explanatory reach of content-based theories are likely to be beyond the reach of STM theories as well. I shall then turn to Stich's first claim, arguing that he has not shown that STM theories will be able to capture behavioral regularities readily explained by content-based theories.

**2. Do STM Theories Capture an Interesting Class of Generalizations Missed by Content-Based Theories?** A substantial part of Stich's book is devoted to establishing that the ascription of belief in folk psychology is both *observer-relative* and irremediably *vague*. The ascription of belief is observer-relative, Stich argues, inasmuch as a subject's belief is iden-

<sup>1</sup>This generalization is the syntactic analogue of the following typical content-based, or RTM, generalization:

For all subjects S, and all sentences in our language P and Q, if S has a belief that can be attributed by a sentence of the form "S believes that p" and if S comes to have a belief that can be attributed by a sentence of the form "S believes that if p then q" then (it is probable that) S will come to have a belief which can be attributed by a sentence of the form "S believes that q" where 'p' is replaced by P and 'q' is replaced by Q throughout. (p. 141)

tified as the belief *that p* just in case it is content similar to the belief that underlies *our own* normal assertion of 'p'. It is vague inasmuch as (i) observers are said to judge content similarity along at least *three* distinct dimensions of similarity and (ii) pragmatic factors (namely, context) determine which of these dimensions plays a decisive role in a given judgment of content similarity. The three dimensions that Stich identifies are "causal-pattern similarity" (patterns of causal interaction with stimuli, behavior, and other mental states; elsewhere called "narrow causal role"), ideological similarity (similarity of doxastic surroundings), and referential similarity.

The observer-relativity of folk psychological ascriptions of belief is alleged to create a special problem for content-based theories, since for each dimension of similarity there are said to be subjects sufficiently different from ourselves as to make it impossible to find a content sentence to describe their beliefs. Since generalizations stated over content sentences will therefore have indeterminate application in these cases, so-called exotic subjects will be beyond the explanatory reach of content-based theories.

Syntactic theories, by contrast, will allegedly have no special difficulty characterizing the mental states of exotic subjects, as they characterize mental states not by content sentences but by the syntactic objects to which the states are mapped. The latter are selected by the theorist "with an eye to giving the simplest and most powerful account of the causal links among stimuli, mental states, and behavior and without any concern for similarities and dissimilarities between the subject and the theorist" (p. 158).

Stich assembles what I take to be a compelling case for the claim that belief ascription in folk psychology is vague and observer-relative. The relevant question, however, is whether ascriptions of content in a developed, scientific psychology will, of necessity, be equally vague and observer-relative.

Stich provides no argument that they will. A scientific psychology might plausibly be expected to reduce the vagueness inherent in our everyday folk notion of content. After all, the folk notion serves many purposes besides the scientific ones of prediction and explanation. Thus, while ascriptions of content in folk psychology may indeed depend pragmatically on three distinct notions of content similarity, it is not obvious that ascriptions of content in a scientific psychology need do so. In a series of recent papers, for example, Jerry Fodor has attempted to articulate a more austere notion of content than that ascribed in folk psychology, namely, so-called *narrow* content.<sup>2</sup> The narrow content of a belief is said to su-

<sup>2</sup>His most recent proposal is in Fodor (1987, chap. 2). See Fodor (1982) for an earlier attempt to define the notion.

pervene on physical and/or functional states of the believer, thus referential similarity plays no role in judgments of narrow content similarity. If a notion of narrow content could be articulated and put to use in an RTM theory, then subjects would not be classed as exotics whose mental states defy subsumption under generalizations defined over contents merely because their beliefs were referentially different from our own.

But whatever the prospects for a notion of content that eliminates the vagueness and observer-relativity that seemingly afflicts folk psychological theories, STM theories seem no better able than content-based theories to characterize adequately the mental states of exotic subjects. Consider Stich's case of Mrs. T, an elderly subject whose memory has deteriorated as a result of degenerative brain disease to the point where her remaining beliefs are no longer ideologically similar to those that we would express using the same content sentences. (Two beliefs are ideologically similar if they are embedded in similar belief networks.) Before the onset of her illness, Mrs. T believed that President McKinley was assassinated. After her memory has deteriorated, she is still disposed to answer "McKinley was assassinated" when asked "What happened to McKinley?", but when asked "Is McKinley dead?" she claims not to know. Clearly, we are disinclined to ascribe to Mrs. T now the belief that she once had, namely, that McKinley was assassinated. For her current belief, whatever it is, is not the belief that would underlie our own sincere utterance of "McKinley was assassinated", because it fails to exhibit the appropriate connections to other beliefs. Indeed, we are at a loss to say what Mrs. T believes now, because we have no content sentence available to characterize her mental state. The problem for content-based theories becomes more acute if we assume that certain of Mrs. T's inferential skills are intact, for example, if told "If McKinley was assassinated then he is buried in Ohio", she will still respond "Then McKinley is buried in Ohio". For now there would seem to be a generalization regarding her behavior, namely, the STM generalization cited earlier, which a content-based theory is unable to capture because it has no way of characterizing the mental state that the generalization subsumes.

Stich claims that an STM theory will have no trouble characterizing Mrs. T's current mental state and subsuming it under the same generalization that described her earlier behavior. I believe this claim to be false. Consider first what the claim that a theory can characterize a mental state amounts to. Clearly, the theory must do more than simply assign a name to the state; the theory must *individuate* the state, that is, it must be able to identify token occurrences of the state over time and across subjects in such a way that the state can be subsumed under lawlike generalizations that predict and explain the behavior of subjects with a good measure of generality. This is precisely what content-based theories seem

unable to do with respect to ideologically exotic subjects. STM theories, Stich argues, *can* so characterize an exotic subject's mental states:

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 a 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. Consider, for example, the case of Mrs. T. 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 [the sample STM generalization cited above], then if the illness simply destroys B-states (or erases mental tokens) without affecting the causal potential of the tokens which remain, the very same generalizations will be true of her after the illness has become quite severe. (p. 158)

While it seems possible that a mental token may simply be erased without affecting the individuation of the remaining states, this is not the correct description of what has happened to Mrs. T. The causal potential of Mrs. T's mental states have clearly changed: the mental state underlying the younger Mrs. T's utterance of "McKinley was assassinated" was potentially connected to behavior and other mental states in countless ways; the mental state underlying the senile Mrs. T's utterance of "McKinley was assassinated", however, is causally *inert*, or nearly so.<sup>3</sup> The latter state can no longer play the role in inference and belief fixation that characterized the former state; counterfactuals true of the earlier state are not true of the latter. It is now false, for example, to say of Mrs. T that she would respond with "McKinley was assassinated" if asked "How did McKinley die?"

The problem for the syntactic theorist is not simply that because the causal potential of the mental state underlying Mrs. T's utterance has changed, syntactic generalizations such as the one cited by Stich which were true of Mrs. T before the onset of her illness will fail to be true of her now. The syntactic theorist cannot even individuate Mrs. T's current mental state as a token of some general state type shared by normal subjects (including Mrs. T before the onset of her illness). All the syntactic theorist can do is designate her state as a token of a new state type virtually unique to Mrs. T. Stich embraces this expediency with alacrity. According to Stich,

neither causal nor ideological distance poses any special problems for an STM theorist. To handle subjects whose basic cognitive processes

<sup>3</sup>A similar point is made by Sterelney (1985) and Marras (1987).

differ from our own, the syntactic theorist may specify a distinct set of wffs (a different 'mental language') and a distinct set of generalizations exploiting the syntactic structure of these wffs. (p. 160)

Consider once again the case of Mrs. T. She is, I have argued, a causal-pattern different subject. After her illness has become acute, a syntactic theorist may simply construct a new syntactic theory and subsume the senile Mrs. T's mental states under the new theory's generalizations. The difficulty with Stich's proposal is precisely that the wffs assigned to Mrs. T's mental states by the new theory don't characterize the mental states of any other subjects, normal or exotic. The syntactic theorist can claim to have characterized Mrs. T's mental states in only a *Pickwickian* sense: he has succeeded in designating them as tokens of certain syntactic types, but they are types that lack all generality. Such "generalizations" as the syntactic theorist can construct for Mrs. T lack all generality: they don't apply across subjects, even across most exotic subjects. They don't even apply to Mrs. T across time, since the causal potential of her mental states is continually changing. They apply only to Mrs. T as she is now.

This is individual psychology with a vengeance. The STM "handles" exotic subjects only insofar as it is willing to abandon our interest in both comparative and developmental psychology. At the very least Stich's proposal to construct a new theory for exotic subjects would make the comparison of exotic subjects and ourselves, or the comparison of different exotics, or the comparison of maturing or aging subjects over time, impossible. We have a strong intuition that our cognitive processes have a lot in common with many subjects who count as exotic, for example, young children and animals. When we ascribe to a dog the belief that the squirrel ran up the tree, we attribute to it a mental state *something like* our own mental state when we believe that the squirrel ran up the tree. We want a theory that not only makes inter-species comparisons possible, but also applies diachronically to members of our own species.<sup>4</sup>

In summary, Stich may be correct in arguing that content theories have a serious problem characterizing the mental states of exotic subjects; however, syntactic theories suffer the same problem. The wffs assigned to

<sup>4</sup>Such developmental theories as we now have suggest that a child undergoes *maturational* changes, that is, the causal potential of the states change over time, yet there are good reasons for saying that the states persist through these changes. For example, work in psycholinguistics suggests that children learn grammar over a period of time. In developing learning theories, we want to be able to say that the rule that the child learns at three years old is the *same* rule that the adult knows, that is, we want to be able to type-identify the state over time, across changes in causal potential. A series of syntactic theories, each individuating mental states by their causal potential, would not enable us to do this.

the senile Mrs. T's mental states characterize them only in a Pickwickian sense—the mental state types they pick out are virtually unique to Mrs. T in her present condition. Stich's proposal to construct new theories to describe the behavior of exotic subjects is a desperate measure—it abandons the project of a developmental or comparative psychology.

**Do STM Theories Miss Important Generalizations Captured by Content-Based Theories?** However poorly content-based theories may fare at characterizing the mental states of exotic subjects, they have been remarkably successful at predicting and explaining the behavior of so-called normal subjects. There remains the question of whether STM theories can duplicate the success of folk psychology in the domain of normal subjects. Both Fodor and Pylyshyn have argued that regularities important for the explanation of behavior can only be captured by generalizations which advert to the content of mental states (Fodor 1981, 1982, 1987; Pylyshyn 1980b, 1984). These generalizations, it is claimed, are beyond the reach of STM theories. Pylyshyn argues as follows:

It simply will not do as an explanation of, say, why Mary came running out of the smoke-filled building, to say that there was a certain sequence of expressions computed in her mind according to certain expression-transforming rules. However true that might be, it fails on a number of counts to provide an explanation of Mary's behavior. It does not show why or how this behavior is related to very similar behavior she would exhibit as a consequence of receiving a phone call in which she heard the utterance "the building is on fire!", or as a consequence of hearing the fire alarm or smelling smoke, or in fact following any event interpretable (given the appropriate beliefs) as generally entailing that the building was on fire. (Pylyshyn 1980b, p. 161)

According to Pylyshyn, the only feature common to the diverse circumstances that would produce running-out-of-the-building behavior on Mary's part is that they give rise to an internal state interpretable as a belief that the building is on fire. The relevant generalization, therefore, can only be captured by appeal to the *content* of Mary's internal states. Syntactic theories, of course, eschew appeals to content, and so, Pylyshyn claims, they will miss a generalization readily captured by content-based theories.

Stich's strategy in response to Pylyshyn's argument is to sketch the explanations that a content theory might give for Mary's behavior and then construct parallel explanations in the STM mold. In version 1 of the content story Mary inhales smoke and is caused to believe that the building is on fire. This belief interacts with the long-standing conditional de-



sire to leave a building if it is on fire to produce the desire to leave the building, which in turn interacts with the belief that if one runs out the door one will leave the building to produce the desire to run out the door. In version 2 of the content story Mary comes to believe that the building is on fire as a result of picking up the phone and hearing "The building is on fire!" From this point the story joins version 1.

Consider Stich's parallel STM explanations. In version 1 of the STM story Mary inhales smoke and is caused to have the B-state F (corresponding to the belief that the building is on fire<sup>5</sup>) by an indirect and complex causal process: inhaling the smoke causes her to have the B-state I (corresponding to the belief that she is inhaling smoke) which interacts with the long-standing B-state  $I \rightarrow N$  (corresponding to the belief that if one is inhaling smoke then there is a fire nearby) which produces in her the B-state N, which together with the long-standing B-state  $N \rightarrow F$  results in her having the B-state F (corresponding to the belief that the building is on fire). The B-state F interacts with the long-standing conditional D-state  $F \rightarrow L$  (corresponding to the desire to leave the building if the building is on fire) to produce the D-state L, which in turn interacts with the B-state  $R \rightarrow L$  (corresponding to the belief that if one runs out the door, then one will leave the building) to produce the D-state R, which eventuates in Mary's running out the door. In version 2 of the STM story, Mary's hearing "The building is on fire!" on the telephone causes her to have the B-state H (corresponding to the belief that she is hearing an utterance of "The building is on fire!") which leads her, via an indirect and complex causal chain, to have the B-state F, and from this point the story joins version 1.

What is the significance of the fact that parallel STM explanations can be given? Stich says:

What the various versions have in common, according to the content-based strong RTM explanation, is that they all lead Mary to *believe that the building is on fire*, and this belief plays an essential role in the etiology of her fleeing behavior. On the purely syntactic explanation, there is a *prima facie* perfect parallel. What the various versions have in common is that they all lead Mary to have the B-state F, and this B-state plays an essential role in her fleeing behavior. (p. 176)

<sup>5</sup>These parenthetical remarks about what B- and D-states correspond to are inserted in Stich's text to make explicit the parallelism between STM and content-based explanations. They do not form any part of the STM explanations themselves. I shall have more to say on this point below.

On both versions of the STM story, Mary's fleeing behavior is caused by (among other things) her being in the B-state F. But this state arises, on the two versions, under diverse physical stimuli (distal *and* proximal). In fact, the causal sequences leading to the B-state F on the two versions of the story are completely different. The STM theorist needs to justify the claim that the two state tokens are *syntactically* type-identical, in light of the fact that they don't play the same causal roles with respect to either stimuli or antecedent mental states. The only *apparent* reason to identify the expressions designated by 'F' in the two versions is that they both *mean* "the building is on fire", but of course the STM theorist can't appeal to meanings in the individuation of mental states.

Stich repeatedly stresses the parallelism between content explanations and his own STM explanations, but the parallelism is not *innocent*. Underlying the STM accounts is the following assumption:

[E]ach of the distinct content sentences used to characterize beliefs and desires in the strong RTM explanation corresponds to a distinct syntactic string. . . . All that is being assumed is a token–token correspondence: each of Mary's beliefs and desires (i.e. each token) corresponds to a token of a syntactic type, and *the syntactic tokens are type distinct when and only when the corresponding belief and desire is accorded a distinct content sentence.* (p. 174, my emphasis)

It is unclear why Stich thinks that the assumption involves only a token–token correspondence. In fact, what is being assumed, as the italicized passage indicates, is a one–one mapping of semantic types onto syntactic types, that is, a type–correspondence between content and syntax. No argument is given for this very strong assumption, and it is, in fact, unlikely to be satisfied, given that types at the two levels of description are to be individuated on independent grounds: syntactic types by the causal roles of mental states in the production of behavior, and semantic types by the various criteria that are involved in content ascription. Since the STM is plausibly construed as abstract neurology, satisfaction of the type–correspondence assumption would require, in effect, that contents map one–one onto equivalence classes of neurological states, such equivalence classes to be defined by the roles of these states in the etiology of behavior. It is the burden of a large part of Stich's book that such a correspondence between the contents of mental states and their causal roles can't be effected; indeed if it could, then Stich's claim that content-based theories miss generalizations that can be captured by syntactic theories would be *patently* false, since both types of theories would capture the same class of generalizations.

The type–correspondence assumption is unsupported, if not simply false. This would seem to leave us with no way of type-identifying the internal

states that give rise to Mary's fleeing behavior under diverse circumstances except by appeal to their *meaning*. In particular, since the states are causally related to different stimuli and other internal states, the STM theorist has no grounds for supposing that they have anything *syntactic* in common. But the STM explanations of Mary's behavior hinge on the theory's ability to type-identify what Stich has called "B-state F" in the two versions, otherwise a counterfactual supporting generalization is lost. The conclusion should be clear. Stich is faced with a dilemma: either the type-correspondence assumption is a reasonable assumption, and, consequently, semantic and syntactic generalizations will be co-extensive, thereby undermining Stich's arguments for preferring a syntactic theory to one which adverts to content; or the assumption is unjustified, in which case Stich has offered no reason to suppose that syntactic theories can capture behavioral generalizations readily captured by content-based theories.

A final point about the parallelism between content and STM explanations: all the examples of STM generalizations in Stich's book are constructed to parallel typical content-based generalizations. One must question whether a syntactic theory could stand on its own, whether it could do any genuine explanatory work. Perhaps the syntactic-based generalizations suggest themselves only where a content story has already been told. Nothing in Stich's book assuages these doubts. Not only does Stich fail to offer any empirical support for the STM, but the book contains not a single example of psychological research modeled on the STM. Stich relies exclusively on artificially constructed analogues of content-based explanations for his few examples of STM generalizations. In the absence of an explanatory practice that conforms to the STM pattern, one must seriously question Stich's claims that the STM provides an adequate, indeed preferred, foundation for psychological theorizing. At the very least, psychological theories constructed in the STM mold seem unable to capture a wide range of folk psychological generalizations without exploiting the explanatory apparatus of folk psychology.

## REFERENCES

- Fodor, J. A. (1981), *Representations*. Cambridge, Mass.: MIT Press.  
 ———. (1982), "Cognitive Science and the Twin-Earth Problem", *Notre Dame Journal of Symbolic Logic* 23: 98–118.  
 ———. (1987), *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, Mass.: MIT Press.  
 Marras, A. (1987), "Critical Notice: Stephen Stich's *From Folk Psychology to Cognitive Science: The Case Against Belief*", *Philosophy of Science* 54: 115–127.  
 Pilyshyn, Z. W. (1980a), "Computation and Cognition: Issues in the Foundation of Cognitive Science", *The Behavioral and Brain Sciences* 3: 111–132.  
 ———. (1980b), "Cognitive Representation and the Process-Architecture Distinction", *The Behavioral and Brain Sciences* 3: 154–167.

- . (1984), *Computation and Cognition: Toward A Foundation for Cognitive Science*. Cambridge, Mass.: MIT Press.
- Sterelny, K. (1985), "Critical Notice: S. P. Stich, *From Folk Psychology to Cognitive Science*", *Australasian Journal of Philosophy* 63: 510–519.
- Stich, S. P. (1983), *From Folk Psychology to Cognitive Science: The Case Against Belief*. Cambridge, Mass.: MIT Press.