On the Type/Token Relation of Mental Representations

Murat Aydede
The University of Chicago
Department of Philosophy
1010 East 59th Street
Chicago, IL 60637, U.S.A.

(773) 702 0513 — Office
(773) 702 8513 — Dept.
(773) 702 9861 — Fax

m-aydede@uchicago.edu
http://humanities.uchicago.edu/faculty/aydede/

1. Introduction

According to the Computational/Representational Theory of Thought (CRTT ≈ Language of Thought Hypothesis, or LOTH), propositional attitudes, such as belief, desire, and the like, are triadic relations among subjects, propositions, and internal mental representations. These representations form a representational *system* physically realized in the brain of sufficiently sophisticated cognitive organisms. Further, this system of representations has a combinatorial syntax and semantics, but the processes that operate on the representations are causally sensitive only to their syntax, not to their semantics. On this approach, a first pass account of propositional attitudes is the following (cf. Field 1978: 37 and Fodor 1987: 17):

CRTT For each propositional attitude A there is a unique and distinct (dedicated)[1] computational/functional relation R, and for all propositions P and all subjects S, S As that P IF, AND ONLY IF, there is a mental representation #P# such that S bears R to #P# and #P# means that P.

So take Jones' belief that snow is white. According to LOTH, Jones stands in a computational relation to a sentence in his LOT (Mentalese) that expresses the proposition that snow is white. If Smith has the same belief, then what is it that Smith and Jones share? According to the

---

[1] This is to convey the basic idea that each type of attitude (e.g., believing) is realized by the same type of computational relation (e.g., being inside the computationally defined B-Box) and by no others. So the mapping from attitudes A into computational relations R is meant to be injective.

formulation just given, they must stand in the same type of computational relation (whatever computational relation underwrites the belief relation, as opposed to, say, the desire relation) to some mental sentence in their respective LOT that means that snow is white. We may equally say that they stand in the same type of computational relation to token sentences in their respective LOT that are of the same semantic type, namely, the type that means that snow is white.

On this view, the typing of token representations (intra- or interpersonally) is exclusively semantic. Call this account the semantic account (SA) of typing Mentalese tokens. Accordingly, two symbol tokens in different heads are of the same type if, and only if, they have the same semantic content. Is this account adequate? Or, do we need more than SA for handling type/token relations in the context of a realist intentional framework? This paper is an investigation into this question and the issues that make the question relevant and important.[2] Obviously, at least two more accounts immediately suggest themselves vis-à-vis typing token representations. So let me briefly describe them before picking up the problems with SA and the need to replace or supplement it with these.

Given that the mental sentences are supposed to be realized by the physical properties of neurons in the brain, an obvious suggestion is to type Mentalese tokens on the basis of such properties. I will call this Physical Account (PA) of typing. On this account, two symbol tokens in different heads are of the same type if, and only if, they have the same physical, or quasi-physical,[3] properties. Even if we leave aside the problem of stating exactly what physical properties of tokens must enter into type-identifying them (since there are obviously so many of them and it is unlikely that all are relevant), PA is committed to a strong form of type-physicalism, i.e. to type-physicalism cast across organisms with respect to propositional attitudes with *particular* content, like the belief *that snow is white*. But how plausible is this? Hard-liners in the physical reductionist or intentional eliminativist camps aside, it seems extremely unlikely that Jones and Smith share the belief that snow is white in virtue of harboring sentence tokens that are physically type-identical. Although no one has proven that type-physicalism of this sort is false, I take it (along with, I hope, almost everyone in the field) that it is extremely unlikely to be true in this form as a doctrine cast across organisms for propositional attitudes with *specific* content, whatever its prospects may be with respect to other psychological states such as feeling

---

[2] I should note that, though I have set up the question in terms of CRTT/LOTH, almost all the issues to be raised would apply to different frameworks, as long as they postulate mental representations of one form or another to underwrite our attributions of mental states. Hence our title.

[3] The reason I say '*quasi*-physical' is that physical properties in this sense can be abstract, higher-order properties. For instance, take a complete specification of the shape of a symbol token. This seems like an abstract property. 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 so, some physical properties of symbols can be multiply realized without being functionally defined or even definable. This is why I call them quasi-physical. Nothing will hang on this however. The physical properties relevant in the individuation of Mentalese tokens are probably all neural properties, which might themselves be functional properties from the perspective of physics.

pain, seeing red, etc. In what follows, I will assume without argument that type-physicalism of this sort is false: I take it that this assumption is not controversial. So I put PA aside for what follows.

Unlike PA, the other remaining option turns out to be popular. We might usefully call it the Narrow Functional Account (NFA) of typing Mentalese tokens: two symbol tokens in different heads are of the same type if and only if they have the same narrow functional role. Functional role is constituted by some of the causal relations of a symbol token, both actual and counterfactual. It is therefore customary to think that the functional role of a symbol token is fixed by the Ramsey sentence of (some of) the law-like generalizations[4] involving the symbol type the token belongs to. This is as it should be, since we want the causal roles of tokens of the same type to come out (more or less) the same *across systems*. This could be secured by assuming that the law-like generalizations in question are interpersonally applicable.

What makes this account narrow is that the generalizations would detail only the causal relations among *proximal* stimuli, other symbol tokens, and a proprietary set of behaviors like basic motor-gestures.[5] The network of causal relations described thereby supervenes on the brain (or central nervous system; at any rate, inside the skin) of organisms. As such, nothing outside the skin (i.e., the physical, social, historical, or cultural environments in which the systems are embedded) is relevant to the type-individuation of symbol tokens. I will take up NFA with more detail in § 3.2 below.

As formulated, SA and CRTT do not say anything about the nature of semantic content. Now this can be conceived either as merely truth-conditionally (broadly) or as involving something other than mere truth-conditions. According to the former, then, the semantic content of a sentence *just* are its truth-conditions (accordingly, the semantic content of a Mentalese term *just* is its denotation or extension).

The latter conception has two standard varieties. First, call a semantic content that is not broad, narrow content. This may be conceived on the model of a Fregean sense (or, modes of presentation), either as publicly accessible (graspable) abstract particulars, or on a more naturalistic level, as conceptual/functional roles of Mentalese sentences or terms, perhaps

---

[4] A cluster theorist would claim that the type-identity of a symbol is given by a substantial majority of the generalizations that would cover its tokens. In what follows, I will generally ignore cluster versions of NFA.

[5] It is crucially important for the functional account to be narrow. For if the functional roles are allowed to extend beyond the skin of organisms, the individuation of symbol tokens cannot be semantic-free; and many proponents of NFA, like Stich (1983), want this account to be semantic-free, for reasons we will discuss later. Devitt (1990) claims that even allowing for proximal stimuli and behavior suffices to make the individuation (narrow) semantic. I think he no longer holds this view. Stich in his celebrated book, *From Folk Psychology to Cognitive Science* (1983), asked the same question I am asking in this paper, and built his argumentative strategy against intentional psychology around it: "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" (43–4). He partitions the space of plausible answers similarly.

explicated as computational roles.[6] The second variety is obtained by amalgamating the broad and narrow contents. On this view, the semantic content of a sentence token is an ordered pair whose elements are the narrow and broad contents as just defined. Call this two-factor content.[7] So what we have here are three sorts of semantic content:

  i.   Broad content (truth-conditions/reference)
  ii.   Narrow content (narrow functional/conceptual roles)
  iii. Two-factor content: the ordered pairs consisting of (i) and (ii), <reference, narrow functional role>

It is of course built into CRTT that propositional attitudes are shared by different subjects and by the same subject at different times. Any account of typing Mentalese tokens or any account of semantic content proposed for CRTT should reveal how publicity is obtained. The above formulation of CRTT says that all that is required is to stand in the same type of computational relation to symbol tokens with the same content without saying what kind of content.

The publicity requirement is certainly not gratuitous. It informs many standard accounts of communication as formation/expression of beliefs and desires through the exercise of the communication media. But more importantly, publicity seems to be the backbone of explanation and prediction of human behavior. It is by attributing propositional attitudes that the folk explain and predict people's behavior (understood to involve verbal behavior). It appears that this activity requires that subjects be subsumed under counterfactual supporting content generalizations that are inter- and intra-personally applicable.[8] Consider some examples:

(a) For any person S, if S desires that P and believes that S can bring it about that P, then S will try to bring it about that P.

---

[6] According to many of its defenders, this notion is supposed to be a semantic notion, partly because the narrow content is standardly (though not universally) conceived as something that *determines* the broad content (relative to a context). It is clear that a Fregean (as a narrow content theorist) who wants to give a naturalistic account of senses by reducing them to functional/computational narrow roles will find herself in a position that is at least extensionally equivalent to NFA. I will come back to this later.

[7] There are niceties of contention among the two-factor theorists that partly stem from the relation between the broad and narrow contents. Those who believe that the two are not independent of each other, perhaps in the sense that the narrow content strongly determines the broad, tend to treat the two factors as a single factor. For instance, on Harman's (1987) version of conceptual role semantics, there is only one factor: the functional role of a symbol conceived as "long-armed", i.e. a role defined on the basis of links extending beyond the skin of organism and reaching out to the denotation/extension of the symbol in the world. In what follows, I will ignore the complications created by this, partly because, I tend to believe, the one- and two-factor theories are equivalent. For an argument to this effect, see Block (1987).

[8] Also, these generalizations are not strict; they are hedged by ceteris paribus clauses.

(b) Thirsty people who believe they are thirsty tend to seek water.
(c) Sons avoid marrying their mothers.
(d) People who believe that x is a dog tend to believe that x is an animal.
(e) People who believe that some woodchucks are nearby tend to take steps to avoid them.
(f) People who believe that Superman is nearby tend to feel safer.
(g) People who believe that a is F and that b is G, sometimes tend to be surprised (amused, worried, shocked, etc.) upon learning that $a=b$. [Consider also cases where $G=\sim F$.]

There are likely to be occasions where these can be used to cover more than one agent (to explain or predict their behavior simultaneously or diachronically), so that the range of the universal quantifier over people may vary from two people picked out in a certain way for the purposes at hand, to entire communities, or even across nations. When they are deployed successfully, these generalizations attribute the *same* propositional attitudes to agents. The agents' behavior is supposed to be explained/predicted by subsuming them under such generalizations. So propositional attitudes attributed to agents to explain/predict their behavior are required to be type-identical across them so they can be subsumed by such generalizations. The success of intentional explanation and prediction seems to depend on this fact.

This is more or less the standard picture of how the folk manage to successfully explain and predict their fellows' behavior. It is assumed by many that this picture will be (or, is already in the process of being) vindicated by the development of a mature cognitive science. There are dissenting voices of course: some would deny that cognitive science is intentional enough (in the required sense) to vindicate folk psychology; some would deny that the folk explain and predict their fellows' behavior by implicitly subsuming them under intentional laws; some others would claim that even if explanation by nomic subsumption is right, it does not require the identity of attitudes attributed.

In what follows I will assume the standard picture without argument. This is not because I think it is unproblematic, but because I want to work out some of its consequences for a correct account of propositional attitudes under CRTT and their publicity.

2. Broad SA and its Discontents

It is a more or less received view that when SA is interpreted as broad (purely denotational/truth-conditional), CRTT will be inadequate. This is primarily because, it is thought, many intentional generalizations like the ones above either turn out to be just false, or do not even make sense, when they are read transparently (i.e., when the operating semantic notions are broad). For instance, (a) and (c) seem to be falsified by Oedipus. He did marry Mom/Jocasta after all, in fact, he *sought* to marry her. That violates (c). Generalization (a) also seems to be falsified by Oedipus' story. Oedipus wished not to marry Mom and believed that he could achieve this, yet he did not avoid marrying her — on the contrary. So Oedipus satisfied the antecedent and flouted the consequent of (a). Similarly, consider a small community whose members, for some reason or other, believe that woodchucks, but not groundhogs, are dangerous ("so described"). So, (e) is falsified by this community on a transparent/broad reading, because they will not always take steps (in relevant ways) to avoid woodchucks/groundhogs whenever they believe they are around. Similarly with (f). And, of course, (g) does not even seem to make much sense on a broad reading.

Although broad content is public, it does not seem to slice mental symbols thinly enough: the examples above seem to show that there is a need to type-distinguish co-extensional symbols. Such observations make a strong case for taking SA to involve narrow or two-factor contents. On the conception of narrow content as functional role, co-extensional symbols can be assigned to distinct semantic types if their functional roles are relevantly different.

There are also the so-called Twin Earth cases. Many think that Putnam's celebrated Twin Earth thought experiment shows, *inter alia*, the untenability of stating psychological generalizations in terms of the broad content of mental states. The typical twin cases involve people with identical physical/functional make-up but different broad contents because, it is claimed, their environments are relevantly different. Twin Earth has XYZ where Earth has H_2O: both substances quench thirst, and they look and taste the same, etc. So, it is claimed, if the psychological generalizations are stated in terms of broad contents, then we are likely to miss important generalizations. For instance, there is something important that is shared by both Twin-Earthians and us when we seek to quench our thirst. If we state the generalization (b) in terms of broad content, then Twin-Earthians will not be subsumed by it. The intuition, however, suggests that with respect to thirst-quenching behavior both Twin-Earthians and we are covered by a single generalization. This intuition is sometimes spelled out by saying that the type-identity of mental symbols must supervene on what is inside the skin (physical as well as functional type-identity) if psychological generalizations are to do their explanatory and predictive job. After all, the attribution of propositional attitudes is supposed to be *causally* explanatory; but broad contents do not supervene on what is inside the skin, so they cannot be causally efficacious vis-à-vis the production of behavior.

What this thought experiment seems to show is that the right psychological taxonomy should supervene on what is inside the skin. What I share with my twin is the narrow functional role of our respective mental tokens; so the mental representations underlying propositional attitudes, according to CRTT, must be typed on the basis of their narrow functional role. Hence, it is claimed, the necessity of narrow content.

Both the Frege cases (the existence of co-extensional but functionally different mental tokens across people) and Twin-Earth cases (the existence of functionally type-identical but extensionally different tokens) strongly motivate the adoption of a CRTT that adverts to narrow or two-factor content. Twin Earth cases, in particular, seem to militate against the relevance of a broad content. In what follows, however, I will generally ignore the impact of Twin-Earth cases, and concentrate on the problems posed by Frege cases. But the results of our discussion of narrow content will bear directly on proposals about how to handle Twin-Earth cases.

3. Narrow Functionalist Accounts

Even though a (narrow) functional role semantics is well motivated as a version of SA which avoids the inadequacies of a purely denotational semantics, it has its own problems, according to some like Fodor, quite fatal problems. Functional roles slice symbols thinly enough, but as a result, functional role semantics loses on publicity; for it is hard to produce interpersonally sharable functional roles for symbols. The issues involved here are difficult and complex. So we had better proceed slowly and cautiously.

In this section, I will try to show two things. First, the project of providing a narrow semantic account for Mentalese on the basis of the tokens' functional roles is the same project as providing a "syntactic" account of typing Mentalese tokens. So the NFA and the narrow content account of typing Mentalese tokens collapse into one (assuming a naturalistic semantics). If one has problems, so does the other. Second, the narrow functional account (in its syntactic or semantic versions) cannot solve the problem of typing Mentalese tokens across people.

Establishing the first point is important. For, on the part of those who reject functional role semantics because of its holistic consequences and embrace broad semantics, as the only semantics needed, there is a tendency to appeal to non-semantic, "syntactic," elements in order to pick up the slack in a purely denotational semantics. Recent Fodor is perhaps the best example of this tendency. I want to show that his position does not fare any better than the position he attacks.

This point will not strike many readers as obvious. So let me elaborate on it briefly. It may be objected, for instance, that whatever problems functionalism faces in *semantics*, a functionalist treatment of Mentalese syntax may be immune to them. I suspect that many people believe it is possible to have it both ways: namely, while semantic functionalism may be problematic in ways Fodor has been pointing out, functional individuation of vehicles *across people* is not.[9] I think this view is mistaken. Seeing why is important not only for understanding Fodor, but also for understanding why semantic individuation is necessary for psychological explanation.

3.1 Syntax and Individuation of Mentalese Tokens

Let us start by a clarificatory/terminological note. Fodor's new position can be summarized as follows: propositional attitudes are three-place relations between a subject, a broad proposition, and a symbol. But there are two ways of typing the symbols: one by broad SA, and where SA is inadequate as in Frege cases, by NFA, which Fodor thinks is exclusively a "syntactic" affair. But what does syntax have to do with NFA? In a passage directly relevant to our concern here, Fodor says, for instance:

> A vehicle is a symbol. A symbol (token) is a spatiotemporal particular which has syntactic and semantic properties and a causal role. Vehicles, like other symbols, are individuated with respect to their syntactic and semantic properties, but *not* with respect to their causal roles. In particular, two vehicle tokens are type distinct if they are syntactically different or if they express different propositions. But type-identical vehicle tokens can differ in their causal roles because the role that a token plays depends not just on which

---

[9] As we will see in a moment, even Fodor himself seems to think that he can have it both ways: in his (1994), he seems to defend a functionalist account of typing Mentalese tokens while at the same time arguing against a functionalist account of their semantics. So he must be thinking that a functionalist individuation of vehicles is immune to the problems of a functionalist semantics. I will come back to his (1994) account below and argue that, if it is meant to be interpersonally applicable, it clashes with his views about functionalist semantics.

type it's a token of, but also on the rest of the world in which its tokening transpires. (This is true of the causal roles of symbols because it's true of the causal roles of everything. Roughly, your causal role depends on what you are, what the local laws are, and what else there is around.) I assume, finally, that vehicles can be type distinct but synonymous; distinct vehicles can express the same proposition. So much for the individuation of vehicles. (1989: 167)

This paragraph is one of those rare occasions where Fodor is explicit about how to type Mentalese symbol tokens. What is especially striking is his apparent eagerness to distinguish "syntactic" individuation from the causal kind. I think it is safe to assume that by 'causal role' Fodor means here something like narrow functional role.[10] This is striking because many people (including Fodor himself at other times)[11] have thought that syntactic individuation of symbols is a species of functional individuation where functional individuation is cashed out, more or less, along the lines provided by NFA. Fodor seems to say here that vehicle tokens are individuated with respect to their syntactic and semantic properties, but not with respect to their functional role. One might reasonably infer from this that individuation of a vehicle with respect to its narrow functional role is *not* syntactic individuation; conversely, the syntactic individuation of a token cannot have much to do with its narrow functional role. Given that syntactic individuation differs from semantic individuation, there seems to be only one option about what syntactic individuation comes to: namely, some form of PA (but see below). Sometimes this is indeed the way syntax is characterized. And it is, I believe, this sense that underlies the following often heard claim:

(S1)   The computational (causal/functional) profile of a symbol is (metaphysically) determined by its syntactic properties or form.

On the other hand, syntactic properties and the syntactic individuation of a token are also often said to be a functional matter:

(S2)   The syntactic type-identity (or, syntactic properties or form) of a symbol is (metaphysically) determined by its computational (causal/functional) profile.

Indeed, Stich, in presenting his celebrated Syntactic Theory of Mind (STM), which is usually considered to be a "de-intentionalized" version of LOTH or CRTT, equates the syntactic type-

---

[10] The rest of the article makes it clear that what Fodor has in mind are the causal interactions of a symbol token *with other tokens*: "A very rough theory of belief individuation might make do with *just* a person, a vehicle, and a content. You get a sharper picture if you allow in a functional role for the vehicle. Loosely speaking, I mean by the functional role of a vehicle the role that it plays in inference; more strictly speaking, I mean its causal role in (certain) mental processes" (1989: 168). See also the rest of his discussion (1989: 168ff.).

[11] For a documentation of Fodor's ambivalence between functional and quasi-physical conceptions of syntax, see Devitt (1991) and my (1993).

identity of a brain sentence token with its narrow functional type-identity: "when [brain] states are viewed as tokens of syntactic types, the [narrow] functional profile exhibited by a [brain] state can be equated with what we have been calling its formal or syntactic properties" (Stich, 1983: 190; cf. 1991: 243–5). In fact, the STM is built upon the premise that it is only when vehicle tokens are individuated with respect to their narrow functional profile across organisms that will we be able to do respectable scientific psychology. As is well known, Stich, unlike Fodor, thinks that SA will prove hopeless in this regard. Hence Stich's proposal for typing Mentalese tokens (across people) is NFA-*cum*-STM.

Obviously, (S1) and (S2) cannot both be true at the same level or in the same sense, since the relation of metaphysical determination is not symmetric. Thus, (S1) and (S2) express at best different conceptions of syntax, or conceptions of syntax at different levels. The issues here are complex, and I would like to spare the reader the taste of sorting them out in this paper.[12] The essential point for our purposes, however, is *not* whether PA or NFA of typing Mentalese tokens is properly called "syntactic." Luckily, we may leave this issue aside here. Our present worry is whether there is a way of (narrow) functionally individuating Mentalese tokens (no matter how it may be called) that is free of the problems attaching to functional individuation of their *semantic content*.

There is, however, an innocent way of understanding the term 'syntax' in this context that makes both PA and NFA syntactic: namely, as Fodor once put it, "being syntactic is a way of not being semantic" (1980a: 227). So any scheme of individuating tokens that does not advert to their semantic properties will count as syntactic. I will have this liberal sense in mind for what follows.

3.2 Narrow Content Functionalism and NFA

According to narrow content functionalism, the content of a Mentalese symbol token is (metaphysically) determined or constituted, at least in part, by *some* of its inferential relations. Getting clear about the relevant notion of inferential relations will make the parallelism transparent. I want to make six points (in no particular order) to clarify this.

First, a content functionalist who aspires to naturalize intentionality, or at least conceives of functionalism as part and parcel of the project of naturalizing semantics, cannot non-circularly appeal to a token's "inferential relations" as such, since the very notion of inference is an intentional one. As noted by Fodor,[13] the usual solution to this problem has been to combine functional role semantics with LOTH or some other sort of computationalism. The idea is that inferential relations are to be cashed out by computational/functional/causal relations among non-intentionally characterized symbol tokens.[14] This point is very important. For it is the first

---

[12] I elaborated my own view of syntax as required by CRTT/LOTH in my (1997a). The interested reader will find there a much lengthier treatment of the issues surrounding syntax, computation, and individuation in the context of a comparison between connectionist and classical LOT models.

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

[14] Although this is the natural and usual solution, I should note that a LOT framework is by no means essential here. Loar (1982) is a good example of a functional role semantics that is

place content functionalism and NFA come into contact, since a content functionalist now needs a non-intentional way of typing symbols so that she can assign semantic content to them on the basis of certain of their computational relations. (I will return to this point in a moment.)

  Second, a symbol token, in a certain obvious sense, has indefinitely many *potential* inferential relations.[15] For instance, #*Fa*#[16] *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 like Modus Ponens, Conjunction Elimination, etc., in her computational repertoire. Call the syntactic ones *L-generalizations*. So, for example, #*Fa*# is inferentially connected to #$\Psi\alpha$#, through Modus Ponens* and the expression #*Fa*$\to\Psi\alpha$# for *any* predicate Ψ and *any* constant α. 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 inferred. No particular conditional need to be mentioned. This makes it clear that the meaning-constitutive inferential relations of a token cannot be the 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.[17] Rather, some of the individuating generalizations must be more restricted, immediate and specific (*S-generalizations*, as we might call them). Such generalizations must not quantify over specific symbols. Their quasi-canonical form can perhaps be given as follows:

(C) For all *S*, if *S* comes to have #*Fx*# in her B-box, then, *ceteris paribus*, *S* will come to have #*Gx*# in her B-box.[18]

---

explicitly non-committal about LOT. See his discussion in (1982: 205–8) and (1983). However, for what follows I will assume the LOT framework for a functional role semantics.

[15] Here I am not distinguishing between subsentential and sentential symbols. Although the notion of 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 determined by those sentences in which they occur.

[16] As previously, in order to avoid long and cumbersome ways of expressing the same thing, I will adopt the following convention: 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/mental expression with a '*' to express whatever its computational/syntactic parallel may be.

[17] For further discussion of the nature and plausibility of L-generalizations as psychological generalizations, see Loar (1982: 71ff.). Loar calls them L-constraints, and he adduces quite a number of interesting and *prima facie* plausible examples.

[18] "B-box" is meant to non-semantically capture whatever computational mechanisms underlie our belief forming and storing capacities. (C) as its stands is not in fact well-formulated: '*x*' in '#*Fx*#' 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-

Here are some candidate 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 some such generalizations (in addition to L-generalizations) that the inferential relations of a token are specified. Notice that such S-generalizations are typically the "de-intentionalized" versions of content generalizations of the sort exhibited above ((b)–(f) in § 1).

   Third, there is the issue of scope. Though 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 extreme, the generalizations may be true only of a single 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 restrict the domain of such generalizations to individuals or even to subcultures or to cultures. Intentional psychology is meant to be the psychology of the folk.[19] Hence content functionalists must intend their S-generalizations to cover a lot of ground.

---

generalizations. This formulation and the following toy examples should be taken at an intuitive level.

[19] Indeed, this was the whole point of Stich's criticism of content-based psychologies: namely, that since content attribution is vague, parochial and ascriber-relative (because it is essentially sensitive to a similarity matrix between *actual* doxastic backgrounds of both the ascriber and the ascribee), content psychologies will not make respectable science for they will leave many doxastically dissimilar organisms outside their scope. His NFA-*cum*-STM is supposed to be superior in this respect since its generalizations, being only syntactic, will not reflect the vagueness, parochialism, and observer-relativity of content ascriptions. This is primarily because, according to Stich, "syntactic" generalizations are indifferent to subjects' *actual* doxastic* background (see Stich 1983: 158–9). But we have already seen that NFA is committed to (content*-specific) S-generalizations; otherwise there is no individuation of tokens, since every token is *potentially* inferentially* connected to every other token. This means that NFA-*cum*-STM style psychologies are essentially committed to casting their generalizations for more or less doxastically* homogeneous populations, since there is a way of seeing S-generalizations

Fourth, as I briefly mentioned in characterizing NFA above, the canonical way of specifying the narrow functional role of a token is via the Ramsey sentence of the generalizations covering that token. Informally, the (narrow) meaning of a token reduces to its functional role. This functional role is fixed on the basis of generalizations that cover it by the standard Ramsey-Lewis-Loar method.[20] However, not all S-generalizations that cover a token need enter into the Ramsey sentence that determines its semantic content. Content functionalism typically assumes that only *some* of these generalizations will contribute to content. Hence, we can talk of the "holist" functional role of a token, with its Ramsey sentence containing *all* the generalizations covering it, and we can also talk of a "localist" functional role with its Ramsey sentence containing only *some* of the generalizations covering the token.[21] The way I initially characterized content functionalism above should be understood in this latter sense. (See below.)

Fifth, since, according to functionalism, the narrow content of a token is meant to be exhaustively 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 a *unique* role for each distinct content. So there must be *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, etc.[22]

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-generalizations (plus L-generalizations) such that *via* Ramsefication they secure a *unique* functional role for any contentful token and this role *is* the (narrow) content of the token. The generalizations of **T** are true of all those with a reasonably similar intentional psychology (more or less assumed to be the Folk sharing a certain conceptual repertoire). So the lawlike generalizations are *interpersonally applicable*.

What is supposed to be wrong with narrow content functionalism so characterized? The short answer is that so characterized it is simply false. There is no Ramsefied theory **T** whose generalizations are (i) lawlike, (ii) determine unique roles, and (iii) whose scope is sufficiently large. To see this at an intuitive level, notice that the functionalist seems to need a distinction

as giving a dispositional description of one's actual belief*-system or doxastic* background. This is ironic, of course. Stich's advertisement to the contrary, his own STM is plagued by problems exactly parallel to those that he claimed plague content-based psychologies. See my (forthcoming.) for a detailed and sustained criticism of Stich in this regard.

[20] See Lewis (1972) and Loar (1982). For an accessible introduction to the method, see Block (1980).

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

[22] Cf. Stich (1983: 27), Loar (1982: 44ff.).

between those S-generalizations that are type-identifying (i.e., that belong to **T**) and those that are not. Intuitively, this would correspond to the analytic/synthetic (a/s) distinction, or perhaps it would reconstruct the a/s distinction in non-intentional terms. What might be the basis of such a distinction? For those like Fodor who think that Quine was right about the non-existence of a principled a/s distinction, there can be no such basis. Fodor seems to take this to show that the functionalist is bound to specify functional roles of tokens in terms of *all* S-generalizations that cover them, and he takes this to show in turn that **T** will differ for each individual.[23] Suppose that, as seems likely, the S-generalization (1) above is true for both of us, but the following is true for you but not for me:

(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#.

In this case, you are covered by a Ramsefied theory that includes (6) as well as (1) whereas I am not. But that would mean that your #bachelor# token and mine are type-distinct since their functional roles are different.[24]

  One way to escape this quandary may be to appeal to lawlikeness. It might be said that (1) and (2) above, and their ilk, are lawlike whereas (4)–(6) are not.[25] But even if lawlikeness can

---

[23] Fodor does not conduct his criticism of content functionalism against the background of clarifications and distinctions I have just given. So it is not always clear how his accusation of holism is supposed to go. He gives what calls the "Ur-argument" that is supposed to show that functional role semantics inevitably leads to holism, but he does not defend its second premise at all, which asserts that no principled distinction can be drawn between those inferential relations that constitute the content of a symbol and those that do not. He simply assumes that Quine was right about the a/s distinction without even bothering to explain what it might come to when translated into a non-intentional idiom as must be done if the functionalist is to avoid circularity. Cf. Fodor (1987: 60ff.), Fodor and Lepore (1992: 23ff.) and Devitt (1996: 10).

[24] In fact, if the theories are different, it is not clear that we can even compare our concepts in this way. In a way, the unit of comparison must be the whole theories, which can perhaps be carried out holistically/topologically if at all; otherwise, they might be said to be incommensurable.

[25] Here is one suggestion: a generalization is *lawlike* if, and only if, in all the nearby nomologically possible worlds, where the antecedent is true of the same population over a given period of time, its consequent is also true. This is weaker than the version where we do not 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. Presumably such a world would not be sufficiently nearby. 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. The emphasis must be on the qualifier 'nearby', which makes the formulation vague enough to accommodate our intuitions.

reconstruct the a/s distinction in non-intentional terms, as some functionalists have acknowledged, there are not enough such S-generalizations ("analytic connections") to secure a unique role for each distinct concept.[26] If so, securing unique roles may require appealing to other S-generalizations, like (4)–(5), thus giving up the lawlikeness, and most likely, interpersonal applicability as well (restricting theories' scopes to individuals or very small group of people). This again seems to imply a dangerous sort holism destructive of a respectable scientific psychology.

It is important to note that the requirements of lawlikeness and publicity pull in opposite directions. To the extent to which a functionalist can furnish interpersonally applicable generalizations and secure a unique role for each symbol type, to that extent she fails the requirement that the generalizations be lawlike. And to the extent that she can give lawlike generalizations and secure unique roles, to that extent she violates the condition that the generalizations be interpersonally applicable. There deos not seem to be an optimal point in the continuum between these extremes such that you can meet both of these requirements *and* secure a unique functional role for each possible symbol type.[27]

If what I have said so far is correct, it becomes relatively easy to see how exactly parallel problems plague NFA. Indeed, the two enterprises are in fact identical as far as S-generalizations are concerned. The reason is this: according to content functionalism, 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 typing the token on the basis of

---

So for instance, it may be said that (1) and (2) are lawlike 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. Consider (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 whether in those worlds, tigers, if they exist, are dangerous. Similarly for many other generalizations. In fact, it appears that we do not 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 the ancient Greeks of this world: they apparently believed that the stars were holes in the celestial spheres that the cosmic fire shows through (cf. Fodor 1987: 88–9). That such generalizations are not lawlike 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 does not seem to be a matter of nomological necessities. There is a certain sense in which the prototype theories 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!

[26] See, for instance, Block (1986: 628–9; 1993: 3–4), Loar (1982: 81ff.).

[27] Curiously, this is more or less acknowledged by leading functional role semanticists (e.g., Block). Hence the destructive holism to which they are said to be committed.

14

its functional role as uniquely specified by S-generalizations. In fact, as I mentioned above in my first clarificatory point about inferential relations, the content-functionalist needs a vocabulary to state the S-generalizations. I noted that this vocabulary cannot be intentional, lest the demands of naturalism be violated. So the vocabulary in stating S-generalizations has 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 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 the relation of believing*, 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 individuated across people on the basis of their non-intentional (and presumably, non-physical) properties that obey a certain set of causal/computational regularities. Such a vocabulary will not generally be available to a functionalist prior to the determination of those regularities on the basis of which she proposes to type Mentalese tokens across systems, i.e. on the basis of their narrow functional profile.[28]

      As we have seen, the trouble with content functionalism arises from the difficulty of finding a robust and stable individuation criterion for functional roles that are interpersonally

---

[28] Brian Loar's (1982) is the most worked out attempt to give a through-going functionalist account of the contents of our beliefs and desires, although his version of functionalism is not narrow in my sense of the term. The procedure I assumed throughout in this paper for specifying the functional roles on the basis of the Ramsey sentence of S-generalizations is a simplified version of his own procedure, which he himself barrows from Ramsey and Lewis, and further develops. For a detailed elaboration of this parallelism, see my (forthcoming). The point I want to make here is that a "syntactic" theorist like Stich is in more serious trouble because he cannot avail himself to intentionally characterized symbol tokens at any stage of theory construction, whereas for a content functionalist like Loar, adverting to the semantic properties of symbols in stating the generalizations *may be* kosher in the beginning: Loar specifies the Ramsey sentence of **T** by reference to the (fine-grained) propositions of beliefs and desires given in his M-constraints (roughly equivalent to our S-generalizations). Once he gets the overall (interpersonally valid) causal structure with the help of such a Ramsey sentence, he proposes an interesting technical procedure to eliminate the reference to propositions. The result looks like Stich's STM. So there may be a way of dealing with getting the right vocabulary to state S-generalizations by *initially* adverting to the semantic properties of symbols. However, if Fodor and I are right about the holistic consequences of content functionalism, a Loar-like procedure, even though it can perhaps solve the problem of supplying the right vocabulary for S-generalizations at the end, cannot avoid a destructive sort of holism. But see Loar (1982) for an argument to the contrary. I am not sure, however, whether he still holds the views developed there. See Schiffer (1987) for a critique of Loar (1982).

applicable. If this is the worry, then the same worry must attach to the project of typing 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 ("syntactic") functionalist à la Stich[29] wants to assign to it the "syntactic" type #water is wet# on exactly the same basis.

It should now be transparent why the job of assigning narrow content to a token on the basis of its narrow functional role is exactly the same as the job of typing 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 amounts to on a functionalist research program.[30]

4. Semantic or "Syntactic"? And Does It Matter?

According to CRTT, as we have seen, belief (as a paradigm propositional attitude) is a three-place relation among an agent, a Mentalese sentence token, and a proposition: the agent who believes that *P* stands in a belief-making computational relation to a sentence token in her LOT that means that *P*. This account is now under pressure to explain what it is for two people to share a belief with a particular content. We have also seen that this demand for publicity is not gratuitous: propositional attitude types are essentially adverted to in psychological explanations, which make heavy use of interpersonally applicable nomic generalizations.

Primarily because of the kind of problems with narrow functionalism that we have discussed above, Fodor, beginning in his (1994),[31] adopted a purely denotational (broad) semantics for propositional attitudes. He also argued for a broad intentional psychology,

---

[29] And Fodor, for that matter, if he wants to use NFA for non-semantically individuating tokens — see below.

[30] It might be thought that the uniqueness problem can be solved by adverting to connections of tokens to proximal inputs and outputs. I argue against this in my (forthcoming). Suffice it to say here that this move is not available to Fodor, since he thinks that there are no lawlike connections between proximal inputs and outputs on the one hand, and concepts*, on the other: "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 do not 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).

[31] Fodor (1989) is where he first argued that propositional attitudes have only broad semantics. But Fodor nevertheless continued to hold (till about 1993) onto a notion of narrow content that did not require individuation by functional role. His notion was an extension of Kaplan's notion of "character" as a function from contexts to denotations/truth-conditions. See my (1997b) for exposition and criticism.

according to which the (nomic) generalizations of scientific psychology advert, or will advert, to broad semantic properties of cognitive states (put differently, the scientific intentional generalizations are to be read transparently). Fodor's position naturally raises the question of how to handle Frege cases, cases where there seems to be a need for type-distinguishing co-extensional but functionally distinct states. In his (1994), Fodor advanced an argument to the effect that Frege cases are to be treated as exceptions (not counterexamples) to broad intentional generalizations. I have argued elsewhere that his argument is demonstrably unsound (invalid *and* at least one of its premises is false).[32] But even if Fodor were right, this would not help with the issue of folk explanations of behavior. For whatever the picture of a scientific intentional psychology Fodor may be envisioning, there is no doubt that the folk routinely advert to opaquely read intentional generalizations (like (b)–(g) above). Fodor may be envisioning a radically revisionistic but still intentional scientific psychology, but as he himself acknowledges, he still needs to give at least a sketch of how a scientific explanation of Frege cases should go. For even if Frege cases are exceptions, they do occur. In fact, if a broad scientific psychology is not to miss any generalizations like (b)–(c), and (e)–(g), and many others like them, we need to be told how such cases are to be handled by a scientific psychology. It is important to note here that it is only when generalizations like (b)–(g) are read opaquely do the folk seem capable of capturing robust cognitive/behavioral regularities not capturable by transparent readings. That is why the folk use them. Furthermore, that the folk do use such opaquely read generalizations heavily and successfully is a datum to be explained (how do they do that?); not something to be explained away.

Fodor's proposal is very sketchy. As I mentioned above, he appeals to the "syntax" of the underlying Mentalese tokens in order to distinguish co-extensional symbols. However, it is not clear what the form of such an explanation is supposed to be. If it is meant to be an instance of an explanation by subsumption under nomic generalizations, then we need to know what kind of generalizations these are. Fodor does not give us any clue here.

In his (1998), Fodor argues that concepts are constituents of propositional attitudes. His account of concepts is atomistic, based as it is on a purely denotational (broad) semantics. But the identity of concepts, on this account, is determined by two factors: a broad *content* assigned to a (Mentalese) *vehicle*. In other words, a concept can be represented by an ordered pair or a 2-tuple: <denotation, vehicle>. So for instance, the concept COW = <cows, #cow#>. Since concepts are constituents of propositional attitudes, the thought that snow is white, say, can be represented by the truth-conditions [snow is white] and the Mentalese sentence #snow is white#: <[snow is white], #snow is white#>. In his (1998) and elsewhere, Fodor is adamant that concepts and propositional attitudes are public: "they're the sorts of things that lots of people can, and do, *share*" (1998: 28).[33] So one might think that the generalizations of an envisioned scientific psychology would advert to these pairs. In fact, it is natural to think this, since, as Fodor says,

---

[32] See Aydede and Robbins (forthcoming).

[33] In my (1998), I argue that Fodor's own account of *concepts* violates his own "non-negotiable" constraint that concepts be public, which he uses in his criticism of non-atomistic accounts of concepts.

> [on] the account I have been considering… [i]f content is broad, then behavior is only determined by content taken *together with modes of presentation*. But if Turing was right about psychological processes being exhaustively syntax driven, it looks, once again, as though *content* per se drops out of psychological explanations. It seems that it's the syntactical properties of modes of presentation that are doing all the work, and the attachment to an *intentional*, as opposed to computational, level of psychological explanation is merely sentimental. (1994: 50)

So one might naturally think that at least in Frege cases an appeal to Mentalese vehicles (modes of presentation) is required anyway.

But then, Fodor is in need of an account of type-individuation of Mentalese vehicles across people. And this has to advert to more than their broad semantic properties, because, as we have seen, it is precisely the inadequacy of a broad semantics that causes the trouble in interpersonal Frege cases.

Can he avail himself to a narrow functional individuation of vehicles across people? On the basis of our discussion so far, I hope it is now clear that he cannot — as long as he thinks that narrow functionalism cannot deliver interpersonally sharable functional roles. For as we have seen, the two projects, narrow content functionalism and a narrow functional individuation of vehicles, are in fact one and the same. As long as there are interpersonally applicable Frege cases, Fodor has no adequate account of what it is for two people to share a propositional attitude.

Interestingly, a two-factor content theorist does not seem any better off. According to such a theorist who accepts CRTT or LOTH, any appeal to underlying Mentalese vehicles, in an account of what it is to share a propositional attitude, will be misguided. Part of the reason for this is that the folk who routinely attribute such states to their fellows seem quite neutral about the syntactic ways of individuating vehicle tokens in question. There is even a question about whether the folk have any idea one way or another about the implementing mechanisms of propositional attitudes, let alone their non-semantic type-identity. Working in a broadly Fregean tradition, a two-factor theorist typically regards the issue of sharing beliefs as a purely semantic matter. For such a theorist, the semantic content of concepts/beliefs may differ even if their extensions/truth-conditions are the same. So Mentalese words/sentences are to be typed only by their two-factor semantic content. As we have seen, the semantic content of a Mentalese token (concept) is represented by an ordered pair, a 2-tuple, whose first element is broad content (denotation/extension), and whose second element is the functional/conceptual role of the token: <denotation, functional role>. Psychological generalizations are intentional in that they advert to such pairs. But a two-factor theorist seems to have the same problem with Frege cases and interpersonal generalizations involving them that Fodor has; namely, that narrow functional roles are not interpersonally shareable.

It seems, then, whether the narrow functional role of a Mentalese vehicle is a factor in a semantic content (and thus is itself semantic), or whether it is a "syntactic" factor in the metaphysical constitution of concepts or thoughts according to CRTT, is largely a verbal issue. The substantive issue concerns the need for an account of what it is to share propositional attitudes such that the psychological generalizations appealed to in intentional

explanations/predictions turn out to be applicable across people whose behavior exhibit commonalities precisely because it is behavior acted out of such shared attitudes. At the moment we seem to lack an account that meets this need.

5. In Lieu of a Conclusion

If our discussion has so far failed to produce a successful account of what it is to share a belief that respects intuitions about intentional explanation, this does not mean that no account can be found. On the contrary, the successful folk practice of explaining/predicting behavior on the basis of generalizations like (a)–(g) suggests that there must be one.

This in turn suggests that in looking for a solution, we should perhaps start with an inquiry into how the folk do it in the first place. And, at least for guidance, a useful place to start may be accounts of propositional attitude ascriptions and their semantics. Indeed, one might quite naturally think that this is the appropriate place to look because after all if the folk are successful in attributing propositional attitudes to their fellows while explaining and predicting their behavior, then the truth-makers for such attributions should be available.

Although this may be the direction in which to go, given the broad range (and not altogether promising history) of attempts to give a semantics of belief reports, and the controversy that surrounds them, it is not clear whether this way will lead to any real advance over what I have covered above.

Interestingly, an increasingly popular trend among semanticists working on belief reports is to appeal to mental representations underlying (realizing/implementing) beliefs. From the perspective of representational realism and CRTT, this is of course welcome. But such accounts seem to be under the same sort of pressure to give an account of what it is that a bunch of people share when someone successfully attributes to them the belief that Superman is around the corner, and explain their behavior on this basis by appealing, say, to the generalization (f) above.

Some recent accounts like Crimmins and Perry (1989), Crimmins (1992), Richard (1990) claim that these people share a mental representation type with a certain broad content, but they do not say much about what it is that makes their token representations belong to the same type in cases where broad content cannot help.[34] When the folk attribute to their fellows (simultaneously or diachronically) "Superman" beliefs (as opposed to "Clark Kent" beliefs), how is it that they succeed in getting at the same Superman representations as being type distinct from Clark Kent representations? Pragmatic considerations, we are told, play a substantial role, as the truth conditions of belief reports will vary according to contexts in which such reports are made. But this is not of much help in cases where we fix the relevant contextual parameters and still end up with the same question unanswered. In a context where (f), as opposed to

(f*) People who believe that Clark Kent is nearby tend to feel safer,

---

[34] Perry and Crimmins talk about beliefs' being cognitive particulars in that they involve concrete mental representations. Unlike types or abstracta, cognitive particulars, as they say, cannot be shared. But the difficulty is precisely how to explain sharing of beliefs involving co-referential representations. This they do not really address.

19

is crucial to the explanation of the collective behavior of a bunch of people (when, say, they run in the direction of where they think Superman is as a result of the perceived danger of something in the opposite direction), we still need to know the following: what makes the attribution of the belief that Superman is around the corner true of each of them in such a way that would make the attribution of the belief that Kent is nearby false and unexplanatory. If the answer to this question involves an appeal to some form of functional role of the mental representations involved, then what we want to know is whether such an appeal is free from the kind of problems discussed above.

I have not argued that any such appeal will inevitably involve these difficulties. But it certainly seems like these semantic accounts of belief reports will be susceptible to the same sort of pressures described in this paper. Of course, this shouldn't come as surprise: on an intentional realist framework, the truth-makers for belief reports are, primarily, the beliefs themselves. So long as we lack an adequate account of what it is to share a belief, we are not likely to have an adequate semantic account of attributions of collective belief.

6. Appendix: A proposal by Fodor

For the sake of completeness, I want to briefly look at a specific proposal made by Fodor for typing Mentalese tokens, which might be (and, in fact, has been) thought to provide an escape from this quandary. In the Appendix of his (1994), keen to avoid type-physicalism, Fodor writes:

> it's an *empirical* issue whether, or to what extent, the type-identity thesis is true, and it's one which RTM ought to leave open at least in principle. In particular, the *coherence* of the type/token distinction for Mentalese oughtn't to rest on the *truth* of the type-identity theory. (1994: 108)

And he goes on to propose a (non-semantic) functional individuation of Mentalese symbol types in the following way:

> I assume that we can recognize the *numerical* identity of the machine's states, hence that the notion of *numerically identical tape states* is available to us to use for the characterization of *type* identity for Mentalese symbols. Suppose that T1 and T2 are token inscriptions that do or can appear on the machine's tapes. Then:
> 1. T1 and T2 are tokens of the same type if, for any machine process (i.e., for any operation compatible with the machine's table), the numerically same output state that was (or would have been) produced by the machine when it was given T1 as input would have been produced if the input had been T2; and vice versa.
> 2. If Ti and Tj belong to the same symbol type by criterion 1, then Tm and Tn belong to the same symbol type if any machine process that yielded (or would have yielded) Ti as output given Tm as input would have issued in Tj as output if the input had been Tn.
> The type/token relation for primitive expressions is closed under 1 and 2. (1994: 108–9)

This is a straightforward functionalist proposal for typing Mentalese tokens.[35] However, it is not clear whether this functionalist position is the same as NFA. On the face of it, it looks very different. For it does not try to specify, in a positive way, a functional role of a symbol token on the basis of anything like S-generalizations covering that token. Instead, it seems to lay down the conditions under which two tokens would be type-identical. These conditions are abstract in that they do not mention any *specific* set of computational/functional roles for tokens to play. The proposal only says that whatever those roles are, if they are identical, i.e. if the tokens are or *would be* treated identically in all computational contexts, then they are of the same type.

Fodor does not say anything as to whether his proposal is meant to apply interpersonally. But from the context of his discussion and the assumptions he makes it seems relatively clear that what he has in mind is essentially a criterion for the *intra*personal type-identity of tokens within a single computational system. Indeed, the assumed numerical identity of tape states does not seem to make sense otherwise.

But perhaps we can extend the proposal in an intuitive way to give a type-identity criterion across different computational systems: two tokens in two different systems are of the same type if the systems are ones which treat or would treat those tokens as identical in all computational contexts (even if we may have no *positive* characterization of this functional role identity). This is, I think, in the spirit of Fodor's original proposal, only less formal.

I do not, however, think that the revised proposal could be rescued by making it formally more rigorous. For even at this intuitive level, the suggestion is question begging at best. Surely, two tokens are of the same type if two machines treat or would treat them identically in all computational contexts. But this is true only if the machines have or would have the same computational architecture, running the same program, and perhaps even having the same background data structures, or background information, depending on how you want to cash out the notion of computational context. Apart from this kind of notion of sameness of machines, we have no independent semantic-free purchase on the notion of identical processing of tokens in all computational contexts by two machines, actual or counterfactual.[36]

---

[35] As he himself says: "If one now thinks about how one would set things up so that a pair of Mentalese tokens might actually satisfy this functionally defined relation of co-typehood, it's very plausible that there would have to be some (perhaps very disjunctive) neurological description, that both would be required to satisfy" (1994: 109).

[36] For exactly this kind of worry, see Fodor (1980b: 106, Reply to Rey), where he writes: "Of course, what we *don't* get is the functional identity of formally identical inscriptions *across* computational systems (specifically, across systems which have different mental operations.) We don't, it seems to me, have an account of what makes arbitrary states of arbitrary computational systems states of the same kind, and I wouldn't expect such an account prior to the elaboration of substantive psychological theories of the systems. I don't think there is any guarantee of getting a general notion of functional identity (since this is, clearly, what is at issue) even *given* such theories. This is one of the places where we really may have to face the possibility of indeterminacy." It is very curious to see Fodor saying this then, since the whole point of "Methodological Solipsism" was that the only scientific psychology we would get is a formal one, where organisms get subsumed by the generalizations of such a psychology in virtue of the

If you assume the identity of machines in this sense, of course, you can have functional role identity of symbols across systems. It is like first postulating that all human brains have exactly the same functional architecture, the same program, and the same background data, and then showing that you can have interpersonal identity criteria for typing tokens of their machine language. But that seems trivial and beside the point.[37]

References

Aydede, Murat (1993). *Syntax, Functionalism, Connectionism, and the Language of Thought,* Ph.D. Dissertation, University of Maryland, College Park.

Aydede, Murat (1997a). "Language of Thought: The Connectionist Contribution," *Minds and Machines*, Vol. 7, No. 1, pp. 1–45.

Aydede, Murat (1997b). "Has Fodor Really Changed his Mind on Narrow Content?" *Mind and Language,* Vol. 12, Nos. 3/4, pp. 422–58.

Aydede, Murat (1998). "Fodor on Concepts and Frege Puzzles," *Pacific Philosophical Quarterly*, Vol. 79, No. 4, pp. 289–294.

Aydede, Murat (forthcoming). "Computation and Functionalism: Can Psychology Be Done 'Syntactically'?" in *Boston Studies in Philosophy of Science* edited by G. Güzeldere and G. Irzık, Dordrecht: Kluwer Academic Publishers.

Aydede, Murat and Philip Robbins (forthcoming). "Are Frege Cases Exceptions to Intentional Generalizations?" in *Canadian Journal of Philosophy*.

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 (1980). "Introduction: What is Functionalism?" in *Readings in Philosophy of Psychology*, Vol. 1, Cambridge, MA: Harvard University Press, 1980.

Block, Ned (1986). "Advertisement for a Semantics for Psychology" in *Midwest Studies in Philosophy*, Vol. X, P.A. French, T.E. Uehling and H.K. Wettstein (eds.), Minneapolis, Minnesota: University of Minnesota Press.

---

formal character of their mental symbols. If formality condition wasn't supposed to apply across systems, it is not clear how Fodor envisioned the generality of a formalistic/rationalist psychology. I guess the answer lies in the fact that at that time, like Stich, he equated typing by narrow functional role with formal typing, and got confused by his talk of the "shape" of inscriptions, a confusion not unlike that of between (S1) and (S2) above. See also Fodor's response to Cummins, Geach, Haugeland and Loar in the same place, where he seems to vacillate between (S1) and (S2). Again, see Devitt (1991) and my (1993) for a documentation of this vacillation.

[37] The last three paragraphs owe a great deal to Charles Wallis' stimulating commentary on a precursor of this paper which I presented at the Central APA meeting in Chicago, April 1996. I would also like to thank Jerry Fodor, Güven Güzeldere, John Kulvicki, Sema Kutay, Jesse Prinz, Philip Robbins, Josef Stern, and Ken Taylor for their helpful comments on earlier versions of this paper.

Block, Ned (1987). "Functional Role and Truth Conditions," *Proceedings of the Aristotelian Society*, LXI, 157–181.

Block, Ned (1993). "Holism, Hyper-analyticity and Hyper-compositionality," *Mind and Language*, Vol. 8, No. 1, pp. 1–26.

Crimmins, Mark and John Perry (1989). "The Prince and the Phone Booth: Reporting Puzzling Beliefs," *Journal of Philosophy*, 86: 685–711.

Crimmins, Mark (1992). *Talk About Beliefs*, Cambridge, MA: MIT Press.

Devitt, Michael (1990). "A Narrow Representational Theory of the Mind" in *Mind and Cognition*, W.G. Lycan (ed.), Oxford, UK: Basil Blackwell, 1990.

Devitt, Michael (1991). "Why Fodor Cannot Have It Both Ways" in *Meaning in Mind: Fodor and his Critics*, B. Loewer and G. Rey (eds.), Oxford, UK: Basil Blackwell, 1991.

Devitt, Michael (1996). *Coming to our Senses: A Naturalistic Program for Semantic Localism*, Cambridge, UK: Cambridge University Press.

Field, Hartry H. (1978). "Mental Representation," *Erkenntnis* 13, 1, pp. 9–61. (Also in *Mental Representation: A Reader*, S.P. Stich and T.A. Warfield (eds.), Oxford, UK: Basil Blackwell, 1994. References in the text are to this edition.)

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, Cambridge, MA: MIT Press, 1981. (Originally appeared in *Behavioral and Brain Sciences* 3:1, 1980.)

Fodor, Jerry A. (1980b). "Author's Response" (Response to peer commentary on "Methodological Solipsism"), *Behavioral and Brain Sciences* 3:1.

Fodor, Jerry A. (1985), "Fodor's Guide to Mental Representation: The Intelligent Auntie's Vade-Mecum," *Mind* 94, 1985, pp. 76–100. (Also in Fodor 1990.)

Fodor, Jerry A. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, Cambridge, MA: MIT Press.

Fodor, Jerry A. (1989). "Substitution Arguments and the Individuation of Belief" in *A Theory of Content and Other Essays*, J. Fodor, Cambridge, MA: MIT Press, 1990. (Originally appeared in *Method, Reason and Language*, G. Boolos (ed.), Cambridge, UK: The Cambridge University Press, 1989.)

Fodor, Jerry A. (1990). "A Theory of Content, II" in *A Theory of Content and Other Essays*, J. Fodor, Cambridge, MA: MIT Press, 1990.

Fodor, Jerry A. (1991). "Replies" (Ch. 15) in *Meaning in Mind: Fodor and his Critics*, B. Loewer and G. Rey (eds.), Oxford, UK: Basil Blackwell, 1991.

Fodor, Jerry A. (1994). *The Elm and the Expert: Mentalese and Its Semantics*, Cambridge, MA: MIT Press.

Fodor, Jerry A. (1998). *Concepts: Where Cognitive Science Went Wrong*. 1996 John Locke Lectures, Oxford, UK: Oxford University 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*, Oxford, UK: Basil Blackwell.

Harman, Gilbert (1987). "(Nonsolipsistic) Conceptual Role Semantics" in *New Directions in Semantics*, edited by Ernest Lepore, London: Academic Press, pp. 55–81.

Lewis, David (1972). "Psychophysical and Theoretical Identifications," *Australasian Journal of Philosophy* 50, pp. 249–258.

Loar, Brian (1982). *Mind and Meaning*, Cambridge, UK: Cambridge University Press.

Loar, Brian (1983). "Must Beliefs Be Sentences?" in *Proceedings of the Philosophy of Science Association for 1982*, Asquith, P. and T. Nickles (eds.), East Lansing, Michigan, 1983.

Richard, Mark (1990). *Propositional Attitudes: An Essay on Thoughts and How We Ascribe Them*, Cambridge, UK: Cambridge University Press.

Schiffer, Stephen (1987). *Remnants of Meaning*, Cambridge, MA: MIT Press.

Stich, Stephen P. (1983). *From Folk Psychology to Cognitive Science: The Case Against Belief*, Cambridge, MA: MIT Press.

Stich, Stephen P. (1991). "Narrow Content Meets Fat Syntax" in *Meaning in Mind: Fodor and his Critics*, B. Loewer and G. Rey (eds.), Oxford, UK: Basil Blackwell, 1991.