


WHY THERE STILL HAS TO BE A LANGUAGE OF THOUGHT

JERRY A. FODOR

"But why", Aunty asks with perceptible asperity, "does it have to be *a language*?" Aunty speaks with the voice of the Establishment, and her intransigence is something awful. She is, however, prepared to make certain concessions in the present case. First, she concedes that there are beliefs and desires and that there is a matter of fact about their intentional contents; there's a matter of fact, that is to say, about which proposition the intentional object of a belief or a desire is. Second, Aunty accepts the coherence of physicalism. It may be that believing and desiring will prove to be states of the brain, and if they do that's OK with Aunty. Third, she is prepared to concede that beliefs and desires have causal roles and that overt behavior is typically the effect of complex interactions among these mental causes. (That Aunty was raised as a strict behaviorist goes without saying. But she hasn't been quite the same since the sixties. Which of us has?) In short, Aunty recognizes that psychological explanations need to postulate a network of causally related intentional states. "But why," she asks with perceptible asperity, "does it have to be *a language*?" Or, to put it more succinctly than Aunty often does, what - over and above mere Intentional Realism - does the Language of Thought Hypothesis buy? That is what this discussion is about.

A prior question: What - over and above mere Intentional Realism - does the language of Thought Hypothesis *claim*? Here, I think, the situation is reasonably clear. To begin with, LOT wants to construe propositional-attitude tokens as relations to symbol tokens. According to standard formulations, to believe that P is to bear a certain relation to a token of a symbol which means that P. (It is generally assumed that tokens of the symbols in question are neural objects, but this assumption won't be urgent in the present discussion.) Now, symbols have intentional contents and their tokens are physical in all the known cases. And - qua physical - symbol tokens are the right sorts of things to exhibit causal roles. So there doesn't seem to be anything that LOT wants to claim *so far* that Aunty needs to feel uptight about. What, then, exactly is the issue?

Here's a way to put it. Practically everybody thinks that the *objects* of intentional states are in some way complex: for example, that what you believe when you believe that John is late for dinner is something composite whose elements are it might be - the concept of John and the concept of being late for dinner (or  as it might be - John himself and the property of being late for dinner). And, similarly, what you believe when you believe that P & Q is also something composite, whose elements are - as it might be - the proposition that P and the proposition that Q.

But the (putative) complexity of the *intentional object* of a mental state does not, of course, entail the

complexity of the mental state itself. It's here that LOT ventures beyond mere Intentional Realism, and it's here that Aunty proposes to get off the bus. LOT claims that *mental* states - and not just their propositional objects - *typically have constituent structure*. So far as I can see, this is the *only* real difference between LOT and the sorts of Intentional Realism that even Aunty admits to be respectable. So a defense of LOT has to be an argument that believing and desiring are typically structured states.

Consider a schematic formulation of LOT that's owing to Stephen Schiffer. There is, in your head, a certain mechanism, *an intention box*. 'To make the exposition easier, I'll assume that every intention is the intention to make some proposition true. So then, here's how it goes in your head, according to this version of LOT, when you intend to make it true that P. What you do is, you put into the intention box a token of a mental symbol that *means* that P. And what the box does is, it churns and gurgles and computes and causes and the outcome is that you behave in a way that (*ceteris paribus*) makes it true that P. So, for example, suppose I intend to raise my left hand (I intend to make true the proposition that I raise my left hand). Then what I do is, I put in my intention box a token of a mental symbol that means 'I raise my left hand.' And then, after suitable churning and gurgling and computing and causing, my left hand goes up. (Or it doesn't, in which case the *ceteris paribus* condition must somehow not have been satisfied.) Much the same story would go for my intending to become the next king of France, only in that case the gurgling and churning would continue appreciably longer.

Now, it's important to see that although this *is going to* be a Language of Thought story, it's not a Language of Thought story yet. For so far all we have is what Intentional Realists qua Intentional Realists (including Aunty qua Aunty) are prepared to admit: viz., that there are mental states that have associated intentional objects (for example, the state of having a symbol that means, 'I raise my left hand' in my intention box) and that these mental states that have associated intentional objects also have causal roles (for example, my being in one of these states causes my left hand to rise). That makes the story a Language of Thought story, and not just an Intentional Realist story, is the idea that these mental states that have content also have syntactic structure - constituent structure in particular - that's appropriate to the content that they have. For example, it's compatible with the story I told above that what I put in the intention box when I intend to raise my left hand *is a rock*; so long as it's a rock that's semantically evaluable. Whereas according to the LOT story, what I put in the intention box has to be something like *a sentence* in the present case, it has to be a formula which contains, *inter alia*, an expression that denotes me and an expression that denotes my left hand.

Similarly, on the merely Intentional Realist story, what I put in the intention box when I intend to make it true that I raise my left hand and hop on my right foot might also be a rock (though not, of course, the same rock, since the intention to raise one's left hand is not the same as the intention to raise one's left hand and hop on one's right foot). Whereas according to the LOT story, if I intend to raise my left hand and hop on my right foot, I must put into the intention box a formula which contains, *inter alia*, a subexpression that means *I raise my left hand* and a subexpression that means *I hop on my right foot*.

So then, according to the LOT story, these semantically evaluable formulas that get put into intention boxes typically contain semantically evaluable subformulas as constituents; moreover, they can *share* the constituents that they contain, since, presumably, the subexpression that denotes 'foot' in 'I raise my left foot' is a token of the same type as the subexpression that denotes 'foot' in 'I raise my right

foot.' (Similarly, *mutatis mutandis*, the 'P' that expresses the proposition P in the formula 'P' is a token of the same type as the 'P' that expresses the proposition P in the formula 'P & O'.) If we wanted to be slightly more precise, we could say that the LOT story amounts to the claims that (1) (some) mental formulas have mental formulas as pans; and (2) the pans are 'transportable': the same parts can appear in lots of mental formulas.

It's important to see - indeed, it generates the issue that this discussion is about - that Intentional Realism doesn't logically require the LOT story; it's no sort of *necessary* truth that only formulas - only things that have syntactic structure - are semantically evaluable. No doubt it's puzzling how a rock (or the state of having a rock in your intention box) could have a propositional object; but then, it's no less puzzling how a formula (or the state of having a formula in your intention box) could have a propositional object. It is, in fact, approximately equally puzzling how *anything* could have a propositional object, which is to say that it's puzzling how Intentional Realism could be true. For better or for worse however, Auntie and I are both assuming that Intentional Realism is true. The ques on we're arguing about isn't, then, whether mental states have a semantics. Roughly, it's whether they have a syntax. Or, if you prefer, it's whether they have a *combinatorial* semantics: the kind of semantics in which there are (relatively) complex expressions whose content is determined, in some regular way, by the content of their (relatively) simple pans.

So here, to recapitulate, is what the argument is about: Everybody thinks that mental states have intentional objects; everybody thinks that the intentional objects of mental states are characteristically complex - in effect, that propositions have pans; everybody thinks that mental states have causal roles; and, for present purposes at least, everybody is a functionalist, which is to say that we all hold that mental states are individuated, at least in part, by reference to their causal powers. (this is, of course, implicit in the talk about 'intention boxes' and the like: To be - metaphorically speaking - in the state of having such-and-such a rock in your intention box is just to be - literally speaking - in a state that is the normal cause of certain sorts of effects and/or the normal effect of certain sorts of causes.) What's at issue, however, is the internal structure of these functionally individuated states. Auntie thinks they have none; only the *intentional objects* of mental states are complex. I think they constitute a language; roughly, the syntactic structure of mental states mirrors the semantic relations among their intentional objects. If it seems to you that this dispute among Intentional Realists is just a domestic squabble, I agree with you. But so was the Trojan War.

In fact, the significance of the issue comes out quite clearly when Auntie turns her hand to cognitive architecture; specifically to the question 'What sorts of relations among mental states should a psychological theory recognize?' It is quite natural, given Auntie's philosophical views, for her to think of the mind as a sort of directed graph; the nodes correspond to semantically evaluable mental states, and the paths correspond to the causal connections among these states. To intend, for example, that P & Q is to be in a state that has a certain pattern of (dispositional) causal relations to the state of intending that P and to the state of intending that Q. (E.g., being in the first state is normally causally sufficient for being in the second and third.) We could diagram this relation in the familiar way illustrated in figure 1.

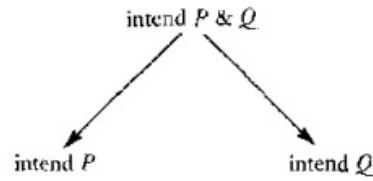


Figure 1

NB: in this sort of architecture, the relation between - as it might be intending that $P \& Q$ and intending that P is a matter of *connectivity* rather than *constituency*. You can see this instantly when you compare what's involved in intending that $P \& Q$ on the LOT story. On the LOT story, intending that $P \& Q$ requires having a sentence in your intention box - or, if you like, in a register or on a tape - one of whose pans is a token of the very same type that's in the intention box when you intend that P , and another of whose pans is a token of the very same type that's in the intention box when you intend that Q .

So, it turns out that the philosophical disagreement about whether there's a Language of Thought corresponds quite closely to the disagreement, current among cognitive scientists, about the appropriate architecture for mental models. If propositional attitudes have internal structure, then we need to acknowledge constituency - as well as causal connectivity - as a fundamental relation among mental states. Analogously, arguments that suggest that mental states have constituent structure *ipso facto* favor Turing/Von Neumann architectures, which can compute in a language whose formulas have transportable parts, as against associative networks, which by definition cannot. It turns out that dear Aunty is, of all things, a New Connectionist Groupie. If she's in trouble, so are they, and for much the same reasons.



Figure 2

In what follows I propose to sketch three reasons for believing that cognitive states - and not just their intentional objects - typically have constituent structure. I don't suppose that these arguments are knockdown; but I do think that, taken together, they ought to convince any Aunty who hasn't *a parti pris*.

First, however, I'd better 'fess up to a metaphysical prejudice that all three arguments assume. I don't believe that there are intentional mechanisms. That is I don't believe that contents *per se* determine causal roles. In consequence, it's got to be possible to tell the whole story about mental causation (the whole story about the implementation of the generalizations that belief/desire psychologies articulate)

without referring to the intentional properties of the mental states that such generalizations subsume. Suppose, in particular, that there is something about their causal roles that requires token mental states to be complex. Then I'm assuming that it does not suffice to satisfy this requirement that these mental states should have *complex intentional objects*.

This is not, by the way, any sort of epiphenomenalism; or if it is, it's patently a harmless sort. There are plenty of cases in the respectable sciences where a law connects a pair of properties, but where the properties that the law connects *don't figure in the story about how the law is implemented*. So, for example, it's a law, more or less, that tall parents have tall children. And there's a pretty neat story about the mechanisms that implement that law. But the property *of being tall* doesn't figure in the story about the implementation; all that figures in that story *is genetic* properties. You got something that looks like figure 2, where the arrows indicate routes of causation.

The moral is that even though it's true that psychological laws generally pick out the mental states that they apply to by specifying the intentional contents of the states, *it doesn't* follow that intentional properties figure in psychological mechanisms. And while I'm prepared to sign on for counterfactual-supporting intentional generalizations, I balk at intentional causation. There are two reasons I can offer to sustain this prejudice (though I suspect that the prejudice goes deeper than the reasons). One of them is technical and the other is metaphysical.

Technical reason: If thoughts have their causal roles in virtue of their contents per se, then two thoughts with identical contents ought to be identical in their causal roles. And we know that this is wrong; we know that causal roles *slice things thinner* than contents do. The thought that P, for example, has the same content as the thought that $\sim \sim P$ on any notion of content that I can imagine defending; but the effects of entertaining these thoughts are nevertheless not guaranteed to be the same. Take a mental life in which the thought that $P \ \& \ (P \ \blacklozenge \ Q)$ immediately and spontaneously gives rise to the thought that Q; there is no guarantee that the thought that $\sim \sim P \ \& \ (P \ \blacklozenge \ Q)$ immediately and spontaneously gives rise to the thought that Q in that mental life.

Metaphysical reason: It looks as though intentional properties essentially involve relations between mental states and *merely* possible contingencies. For example, it's plausible that for a thought to have the content THAT SNOW IS BLACK is for that thought to be related, in a certain way, to the possible (but nonactual) state of affairs in which snow is black; viz., it's for the thought to be true just in case that state of affairs obtains. Correspondingly, what distinguishes the content of the thought that snow is black from the content of the thought that grass is blue is differences among the truth values that these thoughts have in possible but nonactual worlds.

Now, the following metaphysical principle strikes me as plausible: the causal powers of a thing are not affected by its relations to merely possible entities; only relations *to actual* entities affect causal powers. It is, for example, a determinant of my causal powers that I am standing on the brink of a high cliff. But it is not a determinant of my causal powers that I am standing on the brink of a possible-but-nonactual high cliff; I can't throw myself off one *of those*, however hard I try.

Well, if this metaphysical principle is right, and if it's right that intentional properties essentially involve relations to nonactual objects, then it would follow that intentional properties are not per se determinants of causal powers, hence that there are no intentional mechanisms. I admit, however, that

that is a fair number of ifs to hang an intuition on.

OK, now for the arguments that mental states, and not just their intentional objects are structured entities.

1 A methodological argument

I don't, generally speaking, much like methodological arguments; who wants to win by a TKO? But in the present case, it seems to me that Auntie is being a little unreasonable even by her own lights. Here is a plausible rule of nondemonstrative inference that I take her to be at risk of breaking:

Principle P: Suppose there is a kind of event c_1 of which the normal effect

is a kind of event e_1 ; and a kind of event c_2 of which the normal effect is a

kind of event e_2 ; and a kind of event c_3 of which the normal effect is a

complex event $e_1 \& e_2$. Viz.:

$c_1 \rightarrow e_1$

$c_2 \rightarrow e_2$

$c_3 \rightarrow e_1 \& e_2$

Then, *ceteris paribus*, it is reasonable to infer that c_3 is a complex event whose constituents include c_1 and c_2 .

So, for example, suppose there is a kind of event of which the normal effect is a bang and a kind of event of which the normal effect is a stink, and a kind of event of which the normal effect is that kind of a bang and that kind of a stink. Then, according to P, it is *ceteris paribus* reasonable to infer that the third kind of event consists (inter alia) of the co-occurrence of events of the first two kinds.

You may think that this rule is arbitrary, but I think that it isn't; P is just a special case of a general principle which untendentiously requires us to prefer theories that *minimize* accidents. For, if the etiology of events that are e_1 and e_2 does not somehow include the etiology of events that are e_1 but not e_2 , then it must be that there are two ways of producing e_1 events; and the convergence of these (*ex hypothesi*) distinct etiologies upon events of type e_1 is, thus far, unexplained. (It won't do, of course, to reply that the convergence of two etiologies is only a very *little* accident. For in principle, the embarrassment *iterates*. Thus, you can imagine a kind of event c_4 , of which the normal effect is a complex event $e_1 \& e_6 \& e_7$; and a kind of event c_5 , of which the normal effect is a complex event e_1

& e_{10} & e_{12} . . . etc. And now, if P is flouted, we'll have to tolerate a four-way accident. That is, barring P - and all else being equal - we'll have to allow that theories which postulate four kinds of causal histories for e_1 events are just as good as theories which postulate only one kind of causal history for e_1 events. It is, to put it mildly, hard to square this with the idea that we value our theories for the generalizations they articulate.

Well, the moral seems clear enough. Let c_1 be intending to raise your left hand, and e_1 be raising your left hand; let c_2 be intending to hop on your right foot, and e_2 be hopping on your right foot; let c_3 be intending to raise your left hand and hop on your right foot, and e_3 be raising your left hand and hopping on your right foot. Then the choices are: *either we* respect P and hold that events of the c_3 type are complexes which have events of type c_1 as constituents, or we flout P and posit two etiologies for e_1 events, the convergence of these etiologies being, thus far, accidental. I repeat that what's at issue here is the complexity of mental events and not merely the complexity of the propositions that are their intentional objects. P is a principle that constrains etiological inferences, and according to the prejudice previously confessed to - the intentional properties of mental states are *ipso facto* not etiological.

But we're not home yet. There's a way out that Aunty has devised; she is, for all her faults, a devious old dear. Aunty could accept P but deny that (for example) raising your left hand counts *as the same sort* of event on occasions when you just raise your left hand as it does on occasions when you raise your left hand while you hop on your right foot. In effect, Aunty can avoid admitting that *intentions* have constituent structure if she's prepared to deny that *behavior* has constituent structure. A principle like P, which governs the assignment of etiologies to complex events, will be vacuously satisfied in psychology if no behaviors are going to count as complex. But Aunty's back is to the wall; she is, for once, constrained by vulgar fact. Behavior does -very often - exhibit constituent structure, and that it does is vital to its explanation, at least as far as anybody knows: Verbal behavior is the paradigm, of course; everything in linguistics, from phonetics to semantic;s, depends on the fact that verbal forms are put together from recurrent elements; that, for example, [oon] occurs in both 'Moon' and 'June'. But it's not just verbal behavior for whose segmental analysis we have pretty conclusive evidence; indeed, it's not just *human* behavior. It turns out, for one example in a plethora, that bird song is a tidy system of recurrent phrases; we lose 'syntactic' generalizations of some elegance if we refuse to so describe it.

To put the point quite generally, psychologists have a use for the distinction between segmented behaviors and what they call "synergisms." (Synergisms are cases where what appear to be behavioral elements are in fact 'fused' to one another, so that the whole business functions as a unit; as when a well-practiced pianist plays a fluent arpeggio.) Since it's empirically quite clear that not all behavior is synergistic, it follows that Aunty may not, in aid of her philosophical prejudices, simply help herself to the contrary assumption.

Now we are at home. If, as a matter of fact, behavior is often segmented, then principle P requires us to prefer the theory that the causes of behavior are complex over the theory that they aren't, all else being equal. And all else is equal to the best of my knowledge. For if Aunty has any *positive* evidence against the LOT story, she has been keeping it very quiet. Which wouldn't be at all like Aunty, I

assure you.

Argument 2 Psychological processes (why Aunty can't have them for free)

In the cognitive sciences mental symbols are the rage. Psycholinguists in particular, often talk in ways that make Aunty simply livid. For example, they say things like this: "When you understand an utterance of a sentence, what you do is construct *a mental representation* (sic; emphasis mine) of the sentence that is being uttered. To a first approximation, such a representation is a parsing tree; and this parsing tree specifies the constituent structure of the sentence you're hearing, together with the categories to which its constituents belong. Parsing trees are constructed left to right, bottom to top, with restricted look ahead. . ." and so forth depending on the details of the psycholinguist's story. Much the same sort of examples could be culled from the theory of vision (where mental operations are routinely identified with transformations of structural descriptions of scenes) or, indeed, from any other area of recent perceptual psychology.

Philosophical attention is hereby directed to the logical form of such theories. They certainly look to be quantifying over a specified class of mental objects: in the present case, over parsing trees. The usual apparatus of ontological commitment - existential quantifiers, bound variables, and such - is abundantly in evidence. So you might think that Aunty would argue like this: "When I was a girl, ontology was thought to be an a priori science; but now I'm told that view is out of fashion. If, therefore, psychologists say that there are mental representations, then I suppose that there probably are. I therefore subscribe to the Language of Thought hypothesis." That is not, however, the way that Aunty actually does argue. Far from it.

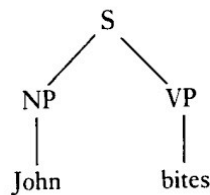


Figure 3

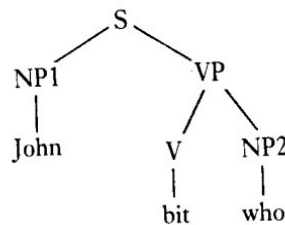


Figure 4

Instead, Aunty regards Cognitive Science in much the same light as Sodom, Gomorrah, and Los Angeles. If there is one thing that Aunty believes in in her bones, it is the ontological promiscuity of psychologists. So in the present case, although psycholinguists may *talk as though* they were:

professionally committed to mental representations, Aunty takes that to be *loose* talk. Strictly speaking, she explains, the offending passages can be translated out with no loss to the explanatory/predictive power of psychological theories. Thus, an ontologically profligate psycholinguist may speak of perceptual processes that construct a parsing tree; say, one that represents a certain utterance as consisting of a noun phrase followed by a verb phrase, as in figure 3.

But Aunty recognizes no such processes and quantifies over no such trees. What she admits instead are (1) the utterance under perceptual analysis (the 'distal' utterance, as I'll henceforth call it) and (2) a mental process which eventuates in the distal utterance being *heard as* consisting of a noun phrase followed by a verb phrase. Notice that this ontologically purified account, though it recognizes mental states with their intentional contents, does not recognize mental representations. Indeed, the point of the proposal is precisely to emphasize as live for Intentional Realists the option of postulating representational mental states and then crying halt. If the translations go through, then the facts which psychologists take to argue for mental representations don't actually do so; and if those facts don't, then maybe nothing does.

Well, but do the translations go through? On my view, the answer is that some do and others don't, and that the ones that don't make the case for a Language of Thought. This will take some sorting out.

Mental representations do two jobs in theories that employ them. First, they provide a canonical notation for specifying the intentional contents of mental states. But second, mental symbols constitute domains over which *mental* processes are defined. If you think of a mental process - extensionally, as it were - as a sequence of mental states each specified with reference to its intentional content, then mental representations provide a mechanism for the construction of these sequences; they allow you to get, in a mechanical way, from one such state to the next by *performing operations on the representations*.

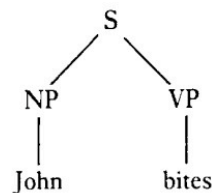


Figure 3

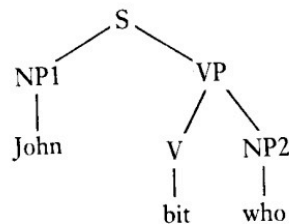


Figure 4

Suppose, for example, that this is how it goes with English *wh*- questions: Such sentences have two constituent structures, one in which the questioned phrase is in the object position, as per figure 4, and

one in which the questioned phrase is in the subject position, as per figure 5. And suppose that the psycholinguistic story is that the perceptual analysis of utterances of such sentences requires the assignment of these constituent structures in, as it might be, reverse order. Well, Aunty can tell *that* story *without* postulating mental representations; *a fortiori* without postulating mental representations that have constituent structure. She does so by talking about *the intentional contents of the hearer's mental* states rather than the mental representations he constructs. "The hearer," Aunty says, "starts out by representing the distal utterance as having 'John' in the subject position and a questioned NP in the object position; and he ends up by representing the distal utterance as having these NPs in the reverse configuration. Thus we see that when it's *properly* construed, claims about 'perceiving as' are all that talk about mental representation ever really comes to." Says Aunty.

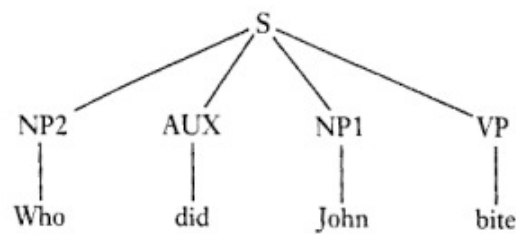


Figure 5

But in saying this, it seems to me that Aunty goes too fast. For what *doesn't* paraphrase out this way is the idea that the hearer gets from one of these representational states to the other *by moving a piece of the parsing tree* (e.g., by moving the piece that represents 'who' as a constituent of the type NP2). This untranslated part of the story isn't, notice, about what intentional contents the hearer entertains or the order in which he entertains them. Rather, it's about the mechanisms that mediate the transitions among his intentional states. Roughly, the story says that the mechanism of mental state transitions *is computational*; and if the story's true, then (a) there must be parsing trees to define the computations over, and (b) these parsing trees need to have a kind of structure that will sustain talk of moving part of a tree while leaving the rest of it alone. In effect, they need to have constituent structure.

I must now report a quirk of Aunty's that I do not fully understand: she refuses to take seriously the ontological commitments of computational theories of mental processes. This is all the more puzzling because Aunty is usually content to play by the following rule: given a well-evidenced empirical theory', either you endorse the entities that it's committed to or you find a paraphrase that preserves the theory while dispensing with the commitments. Aunty holds that this is simply good deportment for a philosopher; and I, for once, agree with her completely. So, as we've seen, Aunty has a proposal for deontologizing the computational story about which state understanding a sentence is: she proposes to translate talk about trees in the head into talk about hearing utterances under descriptions, and that seems to be all right as far as it goes. But it doesn't go far enough, because the ontological commitments of psychological theories are inherited not just from their account of mental states but also from their account of mental processes and the computational account of mental processes would appear to be *ineliminably* committed to mental representations construed as structured objects.

The moral, I suppose, is that if Aunty won't bite the bullet, she will have to pay the piper. As things

stand now, the cost of not having a Language of Thought is not having a theory of thinking. It's a striking fact about the philosophy of mind that we've indulged for the last fifty years or so that it's been quite content to pony up this price. Thus, while an eighteenth-century Empiricist - Hume, say - took it for granted that a theory of cognitive processes (specifically, Associationism) would have to be the cornerstone of psychology, modern philosophers - like Wittgenstein and Ryle and Gibson and Aunty - *have no* theory of thought to speak of. I do think this is appalling; how can you seriously hope for a good account of belief if you have no account of belief fixation? But I don't think it's entirely surprising. Modern philosophers who haven't been overt behaviorists have quite generally been covert behaviorists. And while a behaviorist can recognize mental states - which he identifies with behavioral dispositions - he has literally no use for cognitive processes such as causal trains of thought. The last thing a behaviorist wants is mental causes ontologically distinct from their behavioral effects.

It may be that Aunty has not quite outgrown the behaviorist legacy of her early training (it's painfully obvious that Wittgenstein, Ryle, and Gibson never did). Anyhow, if you ask her what she's prepared to recognize in place of computational mental processes, she unblushingly replies (I quote): "Unknown Neurological Mechanisms." (I think she may have gotten that from John Searle, whose theory of thinking it closely resembles.) If you then ask her whether it's not sort of unreasonable to prefer no psychology of thought to a computational psychology of thought, she affects a glacial silence. Ah well, there's nothing can be done with Aunty when she stands upon her dignity and strikes an Anglo-Saxon attitude - except to try a different line of argument.

Argument 3 Productivity and systematicity

The classical argument that mental states are complex adverts to the productivity of the attitudes. There is a (potentially) infinite set of- for example - belief state types, each with its distinctive intentional object and its distinctive causal role. This is immediately explicable on the assumption that belief states have combinatorial structure; that they are somehow built up out of elements and that the intentional object and causal role of each such state depends on what elements it contains and how they are put together. The LOT story is, of course, a paradigm of this sort of explanation, since it takes believing to involve a relation to a syntactically structured object for which a compositional semantics is assumed.

There is, however, a notorious problem with productivity arguments. The facts of mortality being what they are, not more than a finite part of any mental capacity ever actually gets exploited. So it requires idealization to secure the crucial premise that mental capacities really are productive. It is, for example, quite possible to deny the *productivity* of thought even while admitting that people are forever thinking new things. You can imagine a story - vaguely Gibsonian in spirit - according to which cognitive capacity involves a sort of 'tuning' of the brain. What happens, on this view, is that you have whatever experiences engender such capacities, and the experiences have Unknown Neurological Effects (these Unknown Neurological Effects being mediated, it goes without saying, by the corresponding Unknown Neurological Mechanisms), and the upshot is that you come to have a very large - but finite - number of, as it were, *independent* mental dispositions. E.g., the disposition to think that the cat is on the mat on some occasions; and the disposition to think that 3 is prime on other

occasions; and the disposition to think that secondary qualities are epiphenomenal on other occasions. . . and so forth. New occasions might thus provoke novel thoughts; and yet the capacity to think wouldn't have to be productive. In principle it could turn out, after a lot of thinking, that your experience catches up with your cognitive capacities so that you actually succeed in thinking everything that you are able to. It's no good saying that you take this consequence to be absurd; I agree with you, but Auntie doesn't.

In short, it needs productivity to establish that thoughts have combinatorial structure, and it needs idealization to establish productivity; so it's open to Somebody who doesn't want to admit productivity (because, for example, She doesn't like LOT) simply to refuse to idealize. This is, no doubt, an empirical issue in the very long run. Scientific idealization is demonstrably appropriate if it eventually leads to theories that are independently well confirmed. But vindication in the very long run is a species of cold comfort; perhaps there's a way to get the goodness out of productivity arguments *without* relying on idealizations that are plausibly viewed as tendentious.


Here's how I propose to argue:

a There's a certain property that linguistic capacities have in virtue of the fact that natural languages have a combinatorial semantics.

b Thought has this property too.

c So thought too must have a combinatorial semantics.

Auntie, reading over my shoulder, remarks that this has the form of affirmation of the consequent. So be it; one man's affirmation of the consequent is another man's inference to the best explanation.

The property of linguistic capacities that I have in mind is one that inheres in the ability to understand and produce sentences. That ability is - as I shall say, *systematic*, by which I mean that the ability to produce/understand some of the sentences is *intrinsically* connected to the ability to produce/understand many of the others. You can see the force of this if you compare learning a language the way we really do learn them with learning a language by memorizing an enormous phrase book. The present point isn't that phrase books are finite and can therefore exhaustively describe only nonproductive languages; that's true, but I've sworn off productivity arguments for the duration of this discussion, as explained above. The point that I'm now pushing is that you can learn *any part* of a phrase book *without learning the* rest. Hence, on the phrase book model, it would be perfectly possible to learn that uttering the form of words 'Granny's cat is on Uncle Arthur's mat' is the way to say that Granny's cat is on Uncle Arthur's mat and yet have no idea how to say that it's raining (or, for that matter, how to say that Uncle Arthur's cat is on Granny's mat). I pause to rub this point in. I know  to a first approximation - how to say 'Who does his mother love very much?' in Korean; viz., *ki-iy emma-ka nuku-lil mewu saranna-ci*? But since I did get this from a phrase book, it helps me not at all with saying anything else in Korean. In fact, I don't know how to say anything else in Korean; I have just shot my bolt.

Perhaps it's self evident that the phrase book story must be wrong about language acquisition because a speaker's knowledge of his native language is never like that. You don't, for example, find native speakers who know how to say in English that John loves Mary but don't know how to say in English

that Mary loves John. If you did find someone in such a fix, you'd take that as presumptive evidence that he's not a native English speaker but some sort of a tourist. (This is one important reason why it is so misleading to speak of the block/slab game that Wittgenstein describes in paragraph 2 of the *Investigations* as a "complete primitive language"; to think of languages that way is precisely to miss the systematicity of linguistic capacities - to say nothing of their productivity.)

Notice, by the way, that systematicity (again like productivity) is a property of sentences but not of words. The phrase book model really does fit what it's like to learn the *vocabulary* of English, since when you learn English vocabulary you acquire a lot of basically *independent* dispositions. So you might perfectly well learn that using the form of words 'cat' is the way to refer to cats and yet have no idea that using the form of words 'deciduous conifer' is the way to refer to deciduous conifers. My linguist friends tell me that there are languages - unlike English - in which the lexicon, as well as the syntax, is productive. It's candy from babies to predict that a native speaker's mastery of the vocabulary of such a language is always systematic. Productivity and systematicity run together; if you postulate mechanisms adequate to account for the one, then - assuming you're prepared to idealize - you get the other automatically.

What sort of mechanisms? Well, the alternative to the phrase book story about acquisition depends on the idea, more or less standard in the field since Frege, that the sentences of a natural language have a combinatorial semantics (and, *mutatis mutandis*, that the lexicon does in languages where the lexicon is productive). On this view, learning a language is learning a perfectly general procedure for determining the meaning of a sentence from a specification of its syntactic structure together with the meanings of its lexical elements. Linguistic capacities *can't help but be* systematic on this account, because, give or take a bit the very same combinatorial mechanisms that determine the meaning of any of the sentences determine the meaning of all of the rest.

Notice two things:

First, you can make these points about the systematicity of language without idealizing to astronomical computational capacities. *Productivity* is involved with our ability to understand sentences that are a billion trillion zillion words long. But systematicity involves facts that are much nearer home: such facts as the one we mentioned above, that no native speaker comes to understand the form of words 'John loves Mary' except as he also comes to understand the form of words 'Mary loves John.' In so far as there are 'theory neutral' data to constrain our speculations about language, this surely ought to count as one of them.

Second, if the systematicity of linguistic capacities turns on sentences having a combinatorial semantics, the fact that sentences have a combinatorial semantics turns on their having constituent structure. You can't construct the meaning of an object out of the meanings of its constituents unless it has constituents. The sentences of English wouldn't have a combinatorial semantics if they weren't made out of recurrent words and phrases.

OK, so here's the argument: linguistic capacities are systematic, and that's because sentences have constituent structure. But cognitive capacities are systematic too, and that must be because *thoughts* have constituent structure. But if thoughts have constituent structure, then LOT is true. So I win and Auntie loses. Goody!

I take it that what needs defending here is the idea that cognitive capacities are systematic, not the idea that the systematicity of cognitive capacities implies the combinatorial structure of thoughts. I get the second claim for free for want of an alternative account. So then, how do we know that cognitive capacities are systematic?

A fast argument is that cognitive capacities must be at *least as* systematic as linguistic capacities, since the function of language is to express thought. To understand a sentence is to grasp the thought that its utterance standardly conveys; so it wouldn't be possible that everyone who understands the sentence 'John loves Mary' also understands the sentence 'Mary loves John' if it weren't that everyone who can *think the thought* that John loves Mary can also think the thought that Mary loves John. You can't have it that language expresses thought and that language is systematic unless you also have it that thought is as systematic as language is.

And that is quite sufficiently systematic to embarrass Aunty. For, of course, the systematicity of thought does not follow from what Aunty is prepared to concede: viz., from mere Intentional Realism. If having the thought that John loves Mary is just being in one Unknown But Semantically Evaluable Neurological Condition, and having the thought that Mary loves John is just being in another Unknown But Semantically Evaluable Neurological Condition, then it is - to put it mildly - not obviously why God couldn't have made a creature that's capable of being in one of these Semantically Evaluable Neurological conditions but not in the other, hence a creature that's capable of thinking one of these thoughts but not the other. But if it's compatible with Intentional Realism that God could have made such a creature, then Intentional Realism doesn't explain the systematicity of thought; as we've seen, Intentional Realism is exhausted by the claim that there are Semantically Evaluable Neurological Conditions.

To put it in a nutshell, what you need to explain the systematicity of thought appears to be Intentional Realism plus LOT. LOT says that having a thought is being related to a structured array of representations; and, presumably, to have the thought that John loves Mary is *ipso facto* to have access to the same representations, and the same representational structures, that you need to have the thought that Mary loves John. *So of course* anybody who is in a position to have one of these thoughts is *ipso facto* in a position to have the other. LOT explains the systematicity of thought; mere Intentional Realism doesn't (and neither, for *exactly* the same reasons, does Connectionism). Thus I refute Aunty and her friends!

Four remarks to tidy up:

First: This argument takes it for granted that systematicity is *at least sometimes* a contingent feature of thought; that there are *at least some* cases in which it is logically possible for a creature to be able to entertain one but not the other of two content-related propositions.

I want to remain neutral, however, on the question whether systematicity is always a contingent feature of thought. For example, a philosopher who is committed to a strong 'inferential role' theory of the individuation of the logical concepts might hold that you can't, in principle, think the thought that (P or Q) unless you are able to think the thought that P. (The argument might be that the ability to infer (P or Q) from P is *constitutive of having* the concept of disjunction.) If this claim is right, then - to that extent - you don't need LOT to explain the systematicity of thoughts which contain the concept

ox; it simply *follows from* the fact that you can think that (P or Q) that you can also think that P.

Aunty is, of course, at liberty to try to explain all the facts about the systematicity of thought in this sort of way. I wish her joy of it. It seems to me perfectly clear that there could be creatures whose mental capacities constitute a proper subset of our own; creatures whose mental lives - Viewed from our perspective - appear to contain gaps. If inferential role semantics denies this, then so much the worse for inferential role semantics.

Second: It is, as always, essential not to confuse the properties of the attitudes with the properties of their objects. I suppose that it is necessarily true that the *propositions* are 'systematic'; i.e., that if there is the proposition that John loves Mary, then there is also the proposition that Mary loves John. But that necessity is no use to Aunty, since it doesn't explain the systematicity of our capacity to *grasp* the propositions. What LOT explains - and, to repeat, mere Intentional Realism does not - is a piece of our empirical psychology: the *de facto*, contingent connection between our ability to think one thought and our ability to think another.

Third: Many of Aunty's best friends hold that there is something very special about language; that it is only when we come to explaining linguistic capacities that we need the theoretical apparatus that LOT provides. But in fact, we can kick the ladder away: we don't need the systematicity of language to argue for the systematicity of thought. All we need is that it is on the one hand true, and on the other hand not a *necessary* truth, that whoever is able to think that John loves Mary *is ipso facto* able to think that Mary loves John.

Of course, Aunty has the option of arguing the *empirical* hypothesis that thought is systematic only for creatures that speak a language. But think what it would mean for this to be so. It would have to be quite usual to find, for example, animals capable of learning to respond selectively to a situation such that $a R b$, but quite unable to learn to respond selectively to a situation such that $b R a$ (so that you could teach the beast to choose the picture with the square larger than the triangle, but you couldn't for the life of you teach it to choose the picture with the triangle larger than the square). I am not into rats and pigeons, but I once had a course in Comp Psych, and I'm prepared to assure you that animal minds aren't, in general, like that.

It may be partly a matter of taste whether you take it that the minds of animals are *productive*; but it's about as empirical as anything can be whether they are systematic. And - by and large - they are.

Fourth: Just a little systematicity of thought will do to make things hard for Aunty, since, as previously remarked, mere Intentional Realism is compatible with there being no systematicity of thought at all. And this is just as well, because although we can be sure that thought is somewhat systematic, we can't perhaps be sure of just how systematic it is. The point is that if we are unable to think the thought that P, then I suppose we must also be unable to think the thought that we are unable to think the thought that P. So it's at least arguable that to the extent that our cognitive capacities are not systematic, the fact that they aren't is bound to escape our attention. No doubt this opens up some rather spooky epistemic possibilities; but, as I say, it doesn't matter for the polemical purposes at hand. 'The fact that there are any contingent connections between our capacities for entertaining propositions is remarkable when rightly considered. I know of no account of this fact that isn't tantamount to LOT. And neither does Aunty.'

So we've found at least three reasons for preferring LOT to mere Intentional Realism, and three reasons ought to be enough for anybody's Aunty. But is there any general moral to discern? Maybe there's this one:

If you look at the mind from what has recently become the philosopher's favorite point of view, it's the semantic evaluability of mental states that looms large. What's puzzling about the mind is that anything *physical* could have satisfaction conditions, and the polemics that center around Intentional Realism are the ones that this puzzle generates. On the other hand, if you look at the mind from the cognitive psychologist's viewpoint, the main problems are the ones about mental processes. What puzzles psychologists is belief fixation - and, more generally, the contingent, causal relations that hold among states of mind. The characteristic doctrines of modern cognitive psychology (including, notably, the idea that mental processes are computational) are thus largely motivated by problems about mental causation. Not surprisingly, given this divergence of main concerns, it looks to philosophers as though the computational theory of mind is mostly responsive to technical worries about mechanism and implementation; and it looks to psychologists as though Intentional Realism is mostly responsive to metaphysical and ontological worries about the place of content in the natural order. So, deep down, what philosophers and psychologists really want to say to one another is, "Why do you care so much about that?"

Now as Uncle Hegel used to enjoy pointing out, the trouble with perspectives is that they are, by definition, *partial* points of view; the Real problems are appreciated only when, in the course of the development of the World Spirit, the limits of perspective come to be transcended. Or, to put it less technically, it helps to be able to see the whole elephant. In the present case, I think the whole elephant looks like this: The key to the nature of cognition is that mental processes preserve semantic properties of mental states; trains of thought, for example, are generally truth preserving, so if you start your thinking with true assumptions you will generally arrive at conclusions that are also true. The central problem about the cognitive mind is to understand how this is so. And my point is that neither the metaphysical concerns that motivate Intentional Realists nor the problems about implementation that motivate cognitive psychologists suffice to frame this issue. To see this issue, you have to look at the problems about content and the problems about process *at the same time*. Thus far has the World Spirit progressed.

If Aunty's said it once, she's said it a hundred times: Children should play nicely together and respect each other's points of view. I do think Aunty's right about that.

REFERENCES

Fodor, J. A. (in press) "Information and association," *Notre Dame Journal of Formal Logic*.

Fodor, J. A. (1981) *Representation*, Bradford Books/MIT Press.

Fodor, J. A. (1983) "Reply to Brian Loar's 'Must beliefs be sentences?'," in P. Asquith and T. Nickles (eds), *Proceedings of the Philosophy of Science Association for 1982*, East Lansing, Michigan.

Stich, S. P. (1983) *From Folk Psychology to Cognitive Science*, Bradford Books/MIT Press.