Language, Thought and Compositionality*

JERRY A. FODOR

Introduction

Consider the task under which I labor: This is supposed to be part of a series of papers in the millennial spirit. My charge is to find, somewhere in the philosophical landscape, a problem of whose current status I can give some coherent account, and to point the direction in which it seems to me that further research might usefully proceed. And I'm to try to sound reasonably cheerful and optimistic in the course of doing so. No sooner did I begin to ponder these terms of engagement, than it occurred to me that cheer and optimism aren't really my thing; also that I hadn't heard of a topic in the philosophy of mind (which is the only part of philosophy that even I think that I know anything about) which seems to me to be other than a god-awful mess. It struck me that my best course would be to change my name and go into hiding.

But by dint of unflagging efforts, I have actually found an issue of respectable philosophical provenance, on which it seems to me that there has been some progress; one which, I think, it is becoming possible to contemplate resolving in the foreseeable future, if not by demonstration then still with arguments that an impartial bystander might reasonably find persuasive; at least if he is impartial in my direction. My topic is: Which comes first, the chicken or the egg?

Or, by way of narrowing the field a bit, it's the chicken and egg problem as it arises in the philosophy of mind; namely: 'Which comes, first, thought or language?' Or, by way of narrowing the field still further, 'Which comes

*This is the first in a series of Millennial pieces that have been specially commissioned by *Mind & Language* to appear in this volume and the next. Each contributor has been allowed choice as to topic and approach, being asked only to present a personal view of issues that he or she sees as being important in the present context of cognitive studies.

Professor Fodor's piece will appear later this year in a slightly different version in A. O'Hear, (ed.), *Philosophy at the New Millennium*, Cambridge University Press. This is a volume of Royal Institute of Philosophy lectures given in 1999/2000.

Address for correspondence: Department of Philosophy, Rutgers, The State University of New Jersey, New Brunswick, PO Box 270, New Brunswick, NJ 08903-0270, USA.
Email: fodor@ruccs.rutgers.edu

Mind & Language, Vol. 16 No. 1 February 2001, pp. 1–15.

first *in order of explanation*, the content of thought or the content of language?' (I'll borrow a way of speaking from John Searle and sometimes put this question as: 'Is it thoughts or sentences that have 'content in the first instance' or that have 'underived' content'?)

Two caveats: I'm taking for granted that either sentences mean what they do because they express the thoughts that they do, or vice versa (whatever, exactly, vice versa comes to here.) There are, of course, other positions in logical space; it could be that the content of thought and language both derive from some *third thing*; or that neither thought nor language has any content; or that nothing has any content; or that thought has content in the first instance for thinkers who speak English but language has content in the first instance for thinkers who speak Urdu. And so forth. But none of these options strikes me as attractive and I won't discuss them in what follows.

Also I think we can reasonably expect, having once answered the question about the relative priority of thought and language in the explanation of content, that connected issues about which comes first in the order of ontogenetic or phylogenetic explanation should then fall into place. But I won't argue for that here either.

So now, here's the prospectus: In Part 1, I will tell you why I think we are in a better position to sort out the thought/language issue now than we have been till recently. Roughly, my point will be that we are a little clearer than we used to be about one of the conditions of adequacy that an acceptable theory of content must satisfy. Then, in Part 2, I'll tell you why, now that we are a little clearer about that, the question whether it's thoughts or sentences that have content in the first instance begins to look capable of being solved: There are persuasive empirical reasons suggesting that, although a theory of the content of thought might be able to satisfy the adequacy condition in question, it's unlikely that a theory of the content of language could do so. Which is to say that, as between the two, only thought *has* content strictly speaking. Which is, a fortiori, to say that, as between the two, only thought has content 'in the first instance'.

But I do want to stress, here at the outset, the empiricalness of the reasons I'll have on offer. It has often been supposed that there might be an a priori resolution of the thought/language issue. For example, that there might be a persuasive transcendental argument of more or less the form:

> —Only what can meet the conditions for radical translation (or for radical interpretation, or for being learned by induction from behavioral data; or whatever) can have meaning in the first instance;
> —Only (public) languages can meet such conditions;
> —Therefore, linguistic content must be prior to thought content.

I claim that since their conclusion is empirically false, such arguments must be unsound; but, in the present paper, I'll have nothing else against them. So

if it's your view that philosophy needs a firmer grip on truth than matters of fact can provide, I'm afraid you'll find what follows unconvincing. I apologize in advance; but one of the (many) things that I'm not optimistic about is the philosophical reliability of transcendental arguments.

1. Content

Kierkegaard says somewhere that purity of the heart is to will one thing only. By this (no doubt too stringent) standard, there's long been a lot of *mauvais foi* in the metaphysics of meaning. No doubt, philosophers want a theory of content for its very own sake and quite apart from any other of their problems that it might help them to solve. It would be sort of interesting to know what content is, and where it fits in the natural order; and if it doesn't fit in the natural order, it would be sort of interesting to know that, too. But also, at least since Hume, many philosophers have wanted to know what content is because they've thought knowing that is a prerequisite for constructing a theory of *justification*. Justification is what such philosophers *really* care about; more, even, than they care about purity of the heart.

The putative connection between content and justification has, perhaps, never been formulated very precisely. But the intuition is familiar, so I'll just gesture in its direction. Justification consists in giving reasons. But not *any* reason justifies; only *justified* reasons do. If my claim that P is to justify my claim that Q, then my claim that P must itself be justified (or, at a minimum, justifiable); and its justification *had better not be my claim that Q*. But though this principle seems plausible, and is untendentious in some circles, one sniffs, straight off, the threat of a regress: If every reason needs a reason, how does giving reasons end? Arguably, if we're to block the regress, we need to suppose that at least some claims, at least sometimes, at least prima facie, justify themselves. As Christopher Peacocke rightly says: 'This is actually a form of a classical rationalist principle… to the effect that all a posteriori reason-giving relations rest ultimately on a priori reason-giving relations. (Peacocke, 2000).'

But what would a reason that justifies itself be like? This is where the metaphysics of content is supposed to come floating in, like Lohengrin on his swan, to save the epistemologist's bacon. Actually, there are supposed to be two kinds of self-justifying claims, either of which may serve to avoid the regress. One kind is claims that are true 'solely in virtue of what they mean'; and the other kind is claims that can be 'just seen' to be true. Well, here's the point I've been aiming for: Both the fact, that a claim is true solely in virtue of what it means, and the fact that a claim can be 'just seen' to be true, are supposed to depend on the identity (specifically, on the content) of the concepts that are the claim's constituents. It's (let's say) self-justifying that if John is a bachelor, then John is unmarried. Then anybody who knows what the words 'bachelor' and 'unmarried' mean (or anybody

who knows what the content of the corresponding concepts is; or anybody who 'has' the corresponding concepts; philosophers have a variety of ways of saying this) *thereby* knows that the claim that if John is a bachelor then he is unmarried is self– warranting. Likewise, if it's true that RED is a concept that can be 'just seen' to apply, then (in appropriate circumstances, blah, blah) its appearing to apply is (ceteris paribus, blah, blah) warrant for applying it.

The aspect of this I'd like you please to attend to, is that being the kind of concept that can serve as a constituent of a self-justifying claim is supposed to be *concept-constitutive*. Thus it's constitutive of the concept BACHELOR that *if it's a bachelor, then it's unmarried* is a conceptual truth; and, by a natural extension, it's constitutive of having that concept that one is disposed to draw, and to acquiesce in, tokens of that inference. Likewise, if it's part and parcel of the concept RED that it's 'observational', then having RED is (inter alia) being able (at least some of the time) to apply it just by looking . So the moral is that you can maybe get the theory of content to underwrite the theory of justification; and that this is a consummation devoutly to be wished (ceteris paribus). The way to do so is: you allow epistemic clauses in the definitions of such notions as 'concept individuation' and 'concept possession'.

That is, of course, a kind of foundationalism. Like Descartes' kind, it proposes to block epistemological regresses by endorsing appeals to self-justifying claims, and thereby undertakes an obligation to explain how on earth there could be such things. The difference is that, whereas Descartes thought theology might discharge this burden (via there being innate ideas and God's not being a deceiver), the prevailing view since Hume is that it's the theory of meaning that ought do so. It is to follow *from the theory of meaning* that 'red' is an observation term, and that 'bachelor' implies *unmarried*. If you assume that what follows from the theory of meaning is ipso facto a priori, that explains why such claims don't need reasons to warrant them.

The idea that semantics might underwrite justification, thereby doing for us what God wasn't able to do for Descartes, is most of what is called philosophy's 'linguistic turn.' It's visible, in one form or other, from Hume to Carnap to Quine (inclusive) and from Wittgenstein to Ryle to Davidson to Dummett (likewise inclusive). It is, even now I think, properly stigmatized as the received view. This is not quite to say that everybody, everywhere believes it. Old hands will recall a flurry of interest, in the '60s, largely under Hilary Putnam's tutelage, in 'theoretical' inferences, also known as 'arguments to the best explanation.' The suggestion was that, because it is holistic, this kind of 'a posteriori reason giving', does *not* need to be grounded in 'a priori reason giving'. 'Holism' is what you call circularity if you are commending it (likewise, 'circularity' is what you call holism if you are condemning it). Circles don't stop anywhere, so maybe epistemology can do without foundations if it tolerates a few. And maybe it can tolerate a few if they're big enough. Here too the general picture is familiar, so I'll spare you further exposition.

Suffice it, in any case, that you can't expect semantics to do all that work for you for free. The cost, if you insist that it underwrite your theory of justification, is learning to live with a touch of verificationism about meaning and with a touch of the analytic/synthetic distinction as well; the former because the possession conditions for 'observational' concepts include knowing how to apply them, the latter because 'true in virtue of meaning' is supposed to be what 'analytic' means. Over the years, philosophers have invested quite a lot of sophistication in trying to convince themselves that buying a semantic foundation for epistemology at this price is a good bargain. Many have learned to say 'criterion' and the like in public, and not to blush when they do so. My own view is that it's a doomed strategy: The account of content it requires isn't true; and, anyhow, the kind of justification it can buy you isn't worth having. If my tooth hurts, I want it to be that my behavior justifies (indeed, patently, irresistibly justifies) the dentist's believing that I'm in pain. It is, however, no use his belief being justified because 'toothache behavior' is *what we call* the way I've been behaving. I want it to be justified *because I'm behaving this way because my tooth hurts*. Likewise in more dignified cases; $p<.05$ is what they *call* statistical justification by the data. But what you want is not an experimental hypotheses that is justified by the data by definition; what you want is a hypothesis that is justified by the data *because it's true*. Stipulating a 'level of significance' will never buy you that, of course. Truth requires that the world cooperate.

This does all seem to me to be self-evident, and unavoidable, and why it's so hard to take 'analytic epistemology' seriously. But never mind; I'm prepared to agree that others could think otherwise. I hereby concede that it's possible to believe that grounding justification is worth what theories of meaning would have to pay for it. Or, rather, I hereby concede that it *used to be possible* to believe that. But, to come to the main line of my argument, we now know more than we used to about the adequacy conditions that theories of content must satisfy. And, as it turns out, some of these are *incompatible* with recognizing epistemic constraints on either concept individuation or concept possession. That it does turn out this way is, of course, of considerable interest quite aside from how its doing so bears on the thought and language problem. Quite a lot of modern philosophy has been devoted to the quest for a 'critical' method; one that would dismiss, a priori, metaphysical theses that might otherwise have seemed worth considering. 'When we run over libraries persuaded of these principles, what havoc must we make?' Just like the prospects for a foundationalist philosophy of justification, and for the same reasons, the prospects for a 'critical' philosophy turned first on God's not being a deceiver; and then, when that didn't work, on ideas being copies of impressions; and then, when that didn't work, on content being epistemically constituted. If, as I'm suggesting, that doesn't work either, then probably there *can't be* a critical philosophy. So a venerable tradition comes to an end.

But, though I take these considerations very seriously indeed, it's the

thought–and–language issue that I care about for present purposes. I'm going to argue that the same considerations that guarantee that no epistemic conditions are constitutive of the content of words or concepts, strongly suggest that thoughts must be what have content in the first instance. So the prospectus for what's left of the paper is this: first I'll tell you why content can't be epistemically constituted; and then I'll tell you why the reason that it can't be decides the priority between thought and language.

2. Compositionality

I've suggested that it's possible for rational people to disagree about whether semantic notions like content can underwrite epistemic notions like inference and justification. But here is one thing that we know about content *for sure*: It is compositional in whatever it is underived in. If it's linguistic expressions that have content in the first instance, then the content of a sentence can include only what the content of its constituent expressions (and syntax) contribute; and if it's thoughts that have content in the first instance, then the content of a thought can include only what the content of its constituent concepts (and 'logical' syntax) contribute. And if, as between language and thought, only one of the two has compositional content, then that must be the one whose content is underived. Unlike the thesis that semantics should underwrite justification, the thesis that underived content is compositional is not negotiable. So if, as I'm about to argue, these two theses come into conflict, then it's the first that has to be rejected.

So not–negotiable is compositionality that I'm not even going to tell you what it is. Suffice it that the things that the expression (mutatis mutandis the concept) 'brown cow' applies to are exactly the things to which the expressions 'brown' and 'cow' apply. Likewise, the things that 'brown cow in New Jersey' applies to are exactly the ones that 'brown' 'cow' and 'in New Jersey' apply to. Compositionality is the name of whatever exactly it is that requires this kind of thing to be true in the general case. It says (to repeat) that the semantic value of a thought (/sentence) is inherited from the semantic values of its constituents, together with their arrangement. Whatever, exactly, semantic values are; and whatever, exactly 'determined by' means (both these being matters of some obscurity).

Nobody knows exactly what compositionality demands, but everybody knows why its demands have to be satisfied. Here too the arguments are familiar; and, in my view, they're decisive. Both human thought and human language are, invariably, productive and systematic; and the only way that they could be is by being compositional. (Productivity is the property that a system of representations has if it includes infinitely many syntactically and semantically distinct symbols. Systematicity is the property that a system of representations has (whether or not it is productive) if each of the symbols it contains occurs with the same semantic value as a constituent of many different hosts).

For present purposes, we can collapse systematicity and productivity together,[1] and make the point like this: There are, for better or worse, indefinitely many things that English allows you to say about pigeons and the weather in Manhattan. For example: 'The weather in Manhattan is nice for the pigeons early in the summer' and 'The weather is nice in the early summer in Manhattan if you are a pigeon'…and so forth. English being compositional is what explains why so many of the sentences that you can use to say things about pigeons and the weather in Manhattan, share some or all of their vocabulary. So, in particular, the word 'weather' occurs in many, many of these sentences; and in more or less every case where it does, it contributes the very same semantical property to its sentential host; viz a reference to the weather. Barring idioms and such, this is the general case. A word's occurring in one sentence licenses its occurences in many others, and its semantic contribution is the same in all of them. Were this not so, we couldn't explain the familiar pattern according to which natural languages exhibit open ended clusters of semantically and syntactically related forms. Ditto, mutatis mutandis, for thoughts.[2] So that, in brief, is why compositionality is not negotiable.

We can now begin to put some of the pieces together. Here's what we have so far:

> As between the two, at least one of thought and language must be compositional. If *only* one of them is, then that's the one that has content in the first instance.
>
> In a compositional representational system, whatever belongs to the content of a complex symbol is inherited from the contents of its constituents symbols.

These being so, we can establish the first of the major claims of which I would like to convince you: *None of their epistemic properties are constitutive of the contents either of words or of concepts*. The argument for this is utterly straightforward: *the epistemic properties of symbols do not, in general, compose*. In particular, the epistemic properties of sentences are not, in general, predictable from those of the words that they contain; and the epistemic properties of thoughts are not, in general, predictable from those of the concepts they contain. I'll tell you in a moment why this is so; suffice it, for now, just to argue from examples. I'll take *being observational* (i.e. being the kind of term/concept that can be 'just seen' to apply) as the epistemic property in question; but the argument gen-

---

[1] The difference between them matters primarily in polemical contexts where an argument is required for the compositionality of a representational system that is *not* granted to be productive. (See Fodor and Pylyshyn, 1988).

[2] More precisely, ditto for thoughts assuming that the representational theories of mind is true.

eralizes straightforwardly to cases where the epistemic property that's supposed to be content constitutive involves inferential dispositions.

Here's a first approximation to the argument: Suppose (for reductio) that everyone who has the concept RED SQUARE is ipso facto able to recognize red squares in favorable circumstances. Well, but compositionality says that the content-constitutive properties of concepts are exhaustively inherited from their constituents. So, whoever has these concepts must ipso facto be able to recognize red things and square things in favorable circumstances.

But now it seems that something's wrong. For, as a matter of fact, having the concepts RED and SQUARE is *not* sufficient for being able to recognize red squares in favorable circumstances; not even on the assumption that having the concept RED is sufficient for being able to recognize red things in favorable circumstances and having the concept SQUARE is sufficient for being able to recognize square things in favorable circumstances. This is because it's perfectly possible that the favorable conditions for applying RED SQUARE 'screen off' the favorable conditions for applying SQUARE, so that the circumstances that are favorable for recognizing red things are ipso facto *un*favorable for recognizing square things. If that were so then, a fortiori, the favorable conditions for recognizing RED SQUARE could not be inherited from the favorable conditions for recognizing RED and the favorable conditions for recognizing SQUARE. But, that would be contrary to compositionality, for compositionality requires that complex concepts have *only* such semantic properties as they inherit from their constituents. And compositionality isn't negotiable.

I conclude that the possession conditions for *at least some* concepts (in effect, the complex ones; the ones that have compositional structure) are *not* constituted by recognitional capacities. Arguments similar in spirit apply, mutatis mutandis, to preclude other putative epistemic conditions on the possession of complex concepts.[3]

What I haven't shown, however, is what I promised you that I'd argue for; namely, that *no* concepts are constituted by epistemic possessions conditions. For, the following possibility remains open: Having RED requires being good at recognizing red things; and having SQUARE requires being good at recognizing square things. But having RED SQUARE doesn't require being good at recognizing anything. In effect, according to this account, there can be possession conditions on primitive concepts that they don't transmit to their hosts. So an epistemic semantics might work for the former even if compositionality rules it out for the latter.

But this won't do. In fact, (and this is no small matter) the connection that compositionality imposes on the relations between the possession conditions of constituent concepts and the possession conditions of their hosts goes *in both*

[3] For extensive discussion of these issues, see Fodor (1998), chs. 4 and 5.

directions. That is, compositionality requires not just that having the constituent concepts is sufficient for having a host concept, but also (and even more obviously) that having the host concept is sufficient for having its constituents. Or, to put it slightly differently, compositionality requires that host concepts receive their semantic properties solely from their constituents, *and also that constituent concepts transmit all of their semantic properties to their hosts*. The argument is straight-forward; if constituent concepts didn't transmit their possession conditions to hosts, there would be nothing to stop you from having a host concept *without having its constituents*. Since, as a matter of fact, you don't find people who have RED SQUARE but haven't got RED (or haven't got SQUARE), the possession conditions for host concepts must include the possession conditions for their constituents. But this condition couldn't be met if recognitional capacities are concept constitutive.

You can't have RED SQUARE unless you have the concept RED. It certainly *looks* as though you could have RED SQUARE without having a recognitional capacity for red things at large (in favorable circumstances or otherwise). For example, you might have a concept RED that occurs *only* in host contexts. In that case, you'd be able to think *red square* and *red triangle* but not *red* tout court; a fortiori, you wouldn't be able to recognize red things as such, since that requires being able to think *red*. (That kind of mind wouldn't even be very surprising, since thoughts about red tout court are more abstract than thoughts about red squares.) Well, if this *is* possible, and recognised capacities are content constitutive, then you could have RED SQUARE without meeting the (putative) epistemic conditions on RED.[4] But biconditional compositionality requires that satisfying the possession conditions for RED SQUARE *entails* satisfying the possession conditions for RED. The moral would seem to be that there aren't any epistemic conditions on RED. Since this form of argument is perfectly general, I conclude that there aren't any epistemic constraints on any primitive concepts.

I remarked that trying to exhibit epistemic possession conditions for concepts has been the central project in critical philosophy since Hume. That being so, its worth emphasizing not only that epistemic semantics doesn't work as it were de facto, but that there are principled reasons why nobody should ever have expected it to work. The basic point is this: Epistemic capacities are internally connected to the notion of *a good instance:* Except maybe God,

[4] Or, suppose that typical instances of RED SQUARE aren't, in general, *good* instances of RED (that could even be a law of nature). Then you could even have an *observational* concept RED SQUARE without your concept of RED being observational. This sort of thing happens all the time, cf. red hair and red fire engines (also, mutatis mutandis, pet fish and male nurses.) In such cases, the ability to recognize good instances of the Adjective Noun concept in favorable circumstances does not exploit an ability to recognize good instances of the Adjective concept in favorable circumstances. (Typical instances of MALE NURSE are, in fact, not typical instances of MALE or of NURSE.) That's why you may be able reliably to recognize ANs as such without being able reliably to recognize As as such.

having a concept clearly does not require being able to recognize its instances *come what may*. What's moot is whether it requires being able to recognize *good* instances *in favorable circumstances*.

But whereas 'goodinstancehood' is presumably what's central to epistemology, what matters to semantics is plausibly *instancehood per se*. It's surely at least a necessary condition for an expression having the same content as 'bird' that it apply to all and only birds. An expression which applied to all and only *good instances* of birds would, ipso facto, not *mean* 'bird'; at best, it might mean *good instance of bird*. The reason that there's a clash between compositionlity and epistemic semantics is resident exactly here. For, whereas *instancehood* is compositional, *goodinstancehood is not*. Whatever is an instance of RED SQUARE is ipso facto an instance of RED and of SQUARE. But it is *not* the case that whatever is a *good* instance of RED SQUARE is ipso facto a good instance either of RED or of SQUARE. Good instances of PET FISH are neither good instances of PET nor good instances of FISH.

All that being so, it was one of the worst ideas about the metaphysics of meaning that anybody has ever had that the concept of an instance of a concept should somehow be understood in terms of the concept of a good instance of that concept. I guess Plato was the first philosopher to have this bad idea; but you got it in Wittgenstein, and you still do in the unholy alliance between prototype theories of concepts and connectionist theories of mental architecture. It is, indeed, the currently standard view of conceptual content. But it can't be true since, to repeat, goodinstancehood doesn't compose, and compositionality isn't negotiable.

Compositionality rules out an epistemically based semantics. I've been going on about this not only because it's directly germane to the main topic of whether it's thought or language that comes first, but also because so many philosophers have supposed that semantics *has to be* epistemically based, on pain of losing the connection between meaning and justification, and on pain of abandoning the project of constructing a critical philosophy. Well, so be it; I think the idea that there is an interesting connection between meaning and justification is probably mistaken root and branch; what justifies what depends on how things are connected in the world. As for critical philosophies, I wouldn't buy one at a discount. Probably, if you want to refute somebody's metaphysics, you are going to have to read it first.

I do think, however, that there's a consolation for having to give up on epistemological semantics: We now know, with reasonable certainty, whether it's thought or language that comes first.

3. Thought and Language

Here's the argument: If, as between thought and language, only one of them can plausibly be supposed to be compositional, then that is, ipso facto, the one that comes first in order of the explanation of content; the other has only such

secondary content as it 'derives' from the first. But, as a matter of empirical fact, language is pretty clearly *not* compositional; so it can't have content in the first instance. Q.E.D. It's been a great pleasure being here, and I thank you for your kind attention.

What's that? You wish to hear my empirical premise defended? But I am a philosopher, and hence not in the empirical–premise–defending line of work. Still, I'll give it a try if you insist.

David Pears has a recent book called *Hume's System*. His main point, if I read him right, is that scholars have grossly underestimated the influence of the later Wittgenstein on Hume; and that Hume would have done better still if Wittgenstein's influence had been still greater. According to Pears, 'Hume's method…is to use a kind of introspection which is direct and independent of language…. [But]…is it really likely that anything so intricate [as thought] can be investigated in one mind independently of the language in which it would be passed from that mind to another one?' The question is rhetorical. Pears' 'criticism is that [Hume] exaggerates the perspicuity of the mind and then, when he finds himself forced to pay heed to language, he does little more than glance at it in an oblique and perfunctory way.' Now, comes the bit that I'd like you to think about; Pears offers an argument for claiming that looking closely at language structure is a good way of discovering the structure of thought; namely that 'the linguistic expression of a thought has to transmit it in its entirety and so cannot afford to omit or blur anything that is essential to its structure [i.e. to the structure of the thought that the expression transmits]' (all quotes from p.74).

I have two points to put to you; the first is this: I'm prepared to believe that the function of sentences is primarily to express thoughts; and I am likewise prepared to believe that the content of a sentence is, plus or minus a bit, the thought that it is used to express. Indeed, I am prepared to shout all that that from the roof–tops. I'm therefore also perfectly willing to believe in the heuristic value of drawing inferences from the structure of language to the structure of thought as a research strategy in philosophy. (It couldn't, in any case, be more of a waste of time than the currently fashionable project of trying to infer the structure of thought from pictures of the brain; and it's sure to be a great deal cheaper.) For all that, what Pears claims about how language succeeds in expressing thought is flagrantly and ubiquitously untrue. It is, in fact, *not* the case that 'the linguistic expression of a thought has to transmit it' in some way that requires that the sentence '…not…[to] omit or blur anything that is essential to [the thought's] structure.' To the contrary, language is strikingly elliptical and inexplicit about the thoughts it expresses; though, to be sure, it manages to express them all the same.

The second point is that it couldn't be true that language is strikingly elliptical and inexplict about the thoughts that it expresses if language were compositional in anything like strict detail. For, if it were (and assuming that the content of a sentence is, or is the same as, the content of the correspond-

ing thought) the structure of a sentence would indeed have to be explicit about the structure of the thought it expresses; in particular, the constituents of the sentence would have to correspond in a straightforward way to the thought's constituents. For, if there are constituents of your thought that don't correspond to constituents of the sentence you utter, then since compositionality requires that the content of a thought contain all of the content of its constituents, it must be that there was something in the thought that the sentence left out. So you've said less than you intended. And, likewise, if there's some constituent of the sentence that doesn't correspond to a constituent of the thought, then it must be that there's something in the content of the sentence that isn't in the content of the thought. So you've said more that you intended.

The moral is that if language is compositional (and if what a sentence means is the thought that it expresses) then how a sentence is put together must be very explicit about how the corresponding thought is put together. But, as a simple matter of fact, in the general case, sentences are remarkably *in*explicit with respect to how the thoughts they express are put together. So either the content of the thought is different from the content of the sentence that expresses it, or the sentence isn't compositional. I take it that the first disjunct is preposterous; so I take it that the second disjunct must be true.

Actually, I think this is all pretty untendentious. Nobody doubts that a speaker doesn't generally say everything he's thinking. That's true even if what he's thinking is precisely what he intends to communicate. So, for example, you ask me what's the time, and I say 'it's three o'clock'. This reply is inexplict in two rather different ways. For one thing, the sentence is *syntactically* inexplicit; presumably what I've uttered is an abbreviated form of something like 'it's three o'clock here and now.' Second, although the thought I intended to convey is that it's three o'clock *in the afternoon*, Grice requires me to leave some of that out. That's because, even though you don't know exactly what the time is, you presumably do know what time it is plus or minus twelve hours; and I know that you know that; and you know that I know you know that; and so on. Given all that, communicative efficiency constrains the form of words I use to tell you what you asked. What's obvious in the shared intentional context is generally not something that one bothers to say, even if it is part of what one intends to communicate.

My point, however, is rather different. My point is that a perfectly unelliptical, unmetaphorical, undeictic sentence that is being used to express exactly the thought that it is conventionally used to express, often doesn't express the thought that it would if the sentence were compositional. Either (the typical case; see just above) it vastly underdetermines the right thought; or the thought it determines when compositionally construed isn't in fact, the one that it conventionally expresses. Since it's an open secret among formal semanticists that this sort of thing happens all over the place, I'll give just one example by way of pointing to the sort of phenomena I have in mind.

Russell taught us how to read definite descriptions compositionally; an expression of the form 'the x (Fx)' is satisfied by, and only by, the unique individual that is F. There is, however, a light industry devoted to exhibiting the various ways in which this construal does not work. In particular, 'the book is on the table' does not, usually, convey the thought that there exist exactly one book and exactly one table, and the former is on the latter. What happens, instead, is that the speaker uses the expression when he wants to speak of a certain table or a certain book, each of which he has somehow picked out other than by its being the *unique* x such that (Fx)*;* and, correspondingly, it is those objects that the hearer understands that the speaker intends to communicate a thought about; namely, the thought that the one of them is on the other.

Now, definite descriptions aren't freaks; to the contrary, the point they illustrate is really ubiquitous. If you read a sentence as though it were compositional, then the thought that it *ought* to be conventionally used to express often turns out not to be the one that it *is* conventionally used to express. If I could borrow one of Noam Chomsky's intelligent Martians, I bet what he'd say upon considering the way definite descriptions and lots of other expressions work, is that English just isn't compositional. And I think he'd be right to say that.

Four points about all this, and then I will stop pestering you.

(i) The more or less patent uncompositionality of English isn't something that formal semanticists like to admit. Rather, as far as I can tell, it's the house rule in formal semantics that you hold onto the compositionality of natural language, whatever the cost in face implausibility may be. So, for example, there are semanticists according to whom 'the x (Fx)' actually is compositional in its conventional meaning; it's just that the 'the x' in what we loosely call a definite description is in fact some kind of a singular term; maybe it's a funny looking demonstrative. Why on earth any remotely sane language would use as a demonstrative an expression with the syntactic structure of a quantifier is, however, not explained.

I guess what motivates this sort of move is the reflection that if the semantics of English isn't compositional, and if compositionality is not negotiable, then English hasn't got a semantics. And if English hasn't got a semantics, what exactly is the study of the semantics of English the study of? I think there's a perfectly clear answer to this; namely, it's the study of the (compositional) semantics of thought. In a way, that is bad news; whereas sentences wear their constituent structure more or less on their sleeves, what the compositional determinants of thoughts are, is obscure. (One might take the view that that's part of what makes studying the semantics of thought so interesting; but try selling that to a granting agency.) Also, whereas lots of sentences manifest themselves in behaviors (viz. by getting uttered) the behavioral manifestations of thought are notoriously indirect. And practically everybody in cognitive

science and the philosophy of mind is a behaviorist of one sort or another. (A 'sophisticated behaviorist' to use Dan Dennett's oxymoron.)

(ii) No such objections as I've been urging against the compositionality of language can hold against the compositionality of thought. For, whereas the content of a sentence may be inexplicit with respect to the content of the thought it expresses, a thought can't be inexplicit with respect to its own content; there can't be more—or less—to a thought than there is to its content because a thought just *is* its content. If you put this in the language of a representational theory of mind, it comes out something like: A mental representation is ipso facto compositional with respect to the content that a correct semantics would assign to it. This is not a metaphysical mystery; it just spells out what is implicit in claiming that the content of thoughts, unlike that of sentences, is underived.

(iii) Even if language isn't compositional, it doesn't follow that there's a principled problem about how linguistic communication works. Linguistic communication works by negotiation. Since it's a convention of English that you use 'the x (Fx)' to refer to some F that you have in mind, I know that there has to be such an F if what you said when you uttered a sentence containing this formula is true. It may be—indeed, it typically is—contextually obvious which F you're talking about; and if it's not contextually obvious, I can say so, and you can specify further. Eventually, we'll converge on the same F, and you will then be able to tell me (for example) that that F is G. Your thought that the F is G may well contain constituents (ones that you use to think of the F) of which the sentence that you use to express the thought exhibits no counterparts. That's alright if the shared intentional context allows me to figure out what the unexpressed constituents are. (Indeed, it may be alright even if it doesn't. Communication needn't require that I recognize the whole of the thought that prompted your utterance. Referential consensus as to which F is the F in question suffices for most purposes; I needn't also know how you represent that F in the thought that you intend to communicate.)

(iv) Finally, to reiterate the implications for the thought and language problem (which I'm beginning to be afraid that you're beginning to be afraid that I may have forgotten) here's the syllogism:

> As between thought and language, whichever is compositional is the one that has content in the first instance.
>
> The evidence suggests strongly that language is not compositional.
>
> So, unless the evidence is misleading, it's thought, rather than language, that has content in the first instance.

Which is just what we Cartesians have always supposed. There, that's a cheery, optimistic conclusion, isn't it?

Department of Philosophy
Rutgers University

References

Fodor, J. 1998: *In Critical Condition*. Cambridge, MA: MIT Press.

Fodor, J. and Pylyshyn Z. 1988: Connectionism and cognitive architecture. *Cognition*, 28, 3–71.

Peacocke, C. 2000: Fodor on concepts, philosophical aspects, *Mind and Language*, 15, 327–340.

Pears, D. 1991: *Hume's System*. Oxford University Press.