Why Compositionality Won't Go Away: Reflections on Horwich's 'Deflationary' Theory

Jerry Fodor and Ernie Lepore
Center for Cognitive Science
Rutgers University

Introduction

Compositionality is the idea that the meanings of complex expressions (or concepts) are constructed from the meanings of the less complex expressions (or concepts) that are their constituents.[1] Over the last few years, we have just about convinced ourselves that compositionality is the sovereign test for theories of lexical meaning.[2] So hard is this test to pass, we think, that it filters out practically all of the theories of lexical meaning that are current in either philosophy or cognitive science. Among the casualties are, for example, the theory that lexical meanings are statistical structures (like stereotypes); the theory that the meaning of a word is its use; the theory that knowing the meaning of (at least some) words requires having a recognitional capacity for (at least some) of the things that it applies to; and the theory that knowing the meaning of a word requires knowing criteria for applying it. Indeed, we think that only two theories of the lexicon survive the compositionality constraint: *viz*., the theory that all lexical meanings are primitive and the theory that some lexical meanings are primitive and the rest are definitions. So compositionality does a lot of work in lexical semantics, according to our lights.

Well, so imagine our consternation and surprise when, having just about convinced ourselves of all this, we heard that Paul Horwich has on offer a 'deflationary' account of compositionality, according to which, "...the compositionality of meaning imposes *no constraint at all* on how the meaning properties of words are constituted" (154; our emphasis). Surely, we thought, that can't be right; surely compositionality must rule out at least *some* theories about what word meanings are; for example, the theory that they are rocks, or that they are sparrows or chairs; for how could the meanings of complex expressions be constructed from any of those? What, we wondered, is going on here?

We have arrived at a tentative diagnosis, which is that Horwich fails to enforce several distinctions that turn out to be crucial. For example, sometimes he puts his main conclusion in the way we quoted above: "the compositionality of meaning imposes no constraint at all on how the meaning properties of words are constituted". But sometimes,

[1]We assume, for the present discussion, that words express concepts, and that the content of a word is the content of the concept that it expresses. So we'll move back and forth from talk of words to talk of concepts as convenience of exposition suggests. This is, to be sure, to simplify some complicated matters. But it doesn't affect any of the questions that we disagree with Horwich about.

[2]See Fodor, J. and Lepore E, "Why meaning (probably) isn't conceptual role. Mind and Language, vol. 6, no 4, 1991, 329-343; Fodor, J. and Lepore, E. "The pet fish and the red herring; why concepts aren't prototypes", Cognition, vol. 58, no 2, 1996, 243-276; and papers in Fodor, J. In Critical Condition, MIT Press, 1998.

even on the following page, he puts it like this: “understanding one of one’s own complex expressions (non-idiomatically) is, by definition, nothing over and above understanding its parts and knowing how they are combined” (155). Now, *prima facie*, these wouldn’t seem to be at all the same theses. Whereas the first purports to answer a question about the metaphysics of *meaning, viz.,* ‘What linguistic facts about a complex expression are the supervenience base for its meaning properties?’,[3] the second purports to answer a question about the metaphysics of *understanding*, *viz*., ‘What makes it true of a speaker that he understands an expression in his language?’.[4]

We propose, in what follows, to consider how Horwich’s deflationary account of compositionality fares if the distinction between theories of meaning and theories of understanding is properly attended to. Here’s how we think it all turns out:

--Horwich is right to claim that compositionality is neutral with respect to the metaphysics of understanding expressions when ‘understanding’ refers to a (merely) dispositional state; *but not when it refers to an occurrent state.*

--Horwich is right *strictu dictu* to claim that compositionality is neutral with respect to the character of lexical meanings, but only *strictu dictu.*

--Compositionality taken *together with other constraints that semantic theories are required to satisfy* reduces the options in the theory of lexical meaning to the bare minimum enumerated above.

The first part of the paper is about compositionality and understanding, The second part is about compositionality and lexical meaning.

Part I: Compositionality and Understanding

Here’s one version of what Horwich calls “the basic thesis” of his paper: “...once one has worked out how a certain sentence is constructed from primitive syntactic elements, and provided one knows the meanings of those elements, then, automatically and without further ado, one qualifies as understanding the sentence” (155). We will presently argue that the question whether working out its syntax and its lexical content suffices for understanding a sentence is quite independent of the question whether compositionality

[3]It adds to the confusion that although philosophers sometimes use ‘how are word meanings constituted?’ to ask what are word meanings?, they sometimes use it to ask a quite different question, *viz*., what is it about an expression in virtue of which it means what it does? This last question comes up in, e.g., discussions of the ‘naturalization’ of semantics. Possible answers include: ‘it’s causal relations’; ‘it’s something teleological’; ‘it’s the communicative intentions of speaker/hearers, or the tacit conventions that they adhere to’; etc. We mention this only by way of clearing the air. It’s not what either we or Horwich have in mind in the present discussion.

[4] There are philosophers who hold, as a matter of doctrine, that nothing could be a theory of meaning that isn’t also a theory of understanding. Thus Michael Dummett: "Any theory of meaning which was not, or did not immediately yield, a theory of understanding, would not satisfy the purpose for which, philosophically, we require a theory of meaning." (Dummett, M. "What is a theory of meaning?" in The Seas of Language, Oxford University Press, Oxford, 1993, p.4. Perhaps Horwich accepts this view; he doesn’t say.

constrains theories of lexical meaning. But, for the moment, we proceed to consider Horwich's account of understanding in its own right.

We think there is a (perhaps slightly forced) sense in which grasping its syntax and lexicon are indeed sufficient for understand the meaning of a sentence or other complex expressions. But we also think that there's another sense in which it's pretty clearly not. For compositionality implies that, if you are given the syntax and the lexical content of an expression, you have all the information that's relevant to what it means, hence everything you need to understand it. (Maybe there are still things that you need to know about the world; e.g., what in the world the demonstratives demonstrate. But this isn't the sort of issue that either we or Horwich are concerned with.) And we think that there is a sense of 'understanding an expression' in which having all the information that's relevant to understanding it *is* understanding. This is the notion of understanding that is in play when, for example, linguists say that, *qua* English speaker, you here and now understand infinitely many English sentences, including infinitely many whose tokens you have never encountered and never will.

However, there is also a perfectly natural sense of understanding an expression in which you can *fail* to understand one when you encounter it, *even though* you know the language that it belongs to. It's in this sense of 'understanding an expression' that one may still have some figuring out to do even after having grasped the linguistic properties on which the meaning of the expression supervenes. To see that such a situation can arise, consider sentence S, about whose syntax and lexical inventory we will now tell you the complete and unvarnished truth:

S: 'Dogs dogs dogs dog dog dogs.'

Lexicon:

'dog_N' means dog
'-s' means plural
'dog_V' means to dog

Syntax:

$[[Dogs_N\ [dogs_{N1}\ [dogs_{N2}\ dog_{V2}]_{NP}]_{NP}]]_{NP}\ [dog_{V1}\ [dogs_{N3}]]_{VP}]_S$

Clearly, someone might know everything we've just specified about its lexicon and syntax and nevertheless not understand S. Horwich discusses this sort of case, but only very, very briefly. He says, "the length and complexity of expressions whose structures we are able to discern are constrained by psychological factors" (167). These, however, are constraints on our "ability to understand ... words and appreciate how they are combined; but the compositionality of meaning is not amongst those conditions." As far as we can make out, the idea here is that it's stuff *about our psychology* that explains our problems with S, *therefore it is not* facts about S's compositional structure. But if that is the intended argument, the premise clearly doesn't warrant the conclusion. Consider the following Silly Argument:

Silly Argument: It's stuff about your muscles that explains why you can't lift this rock, *therefore it's not* stuff about what the rock weighs.

Surely, the right answer to the Silly Argument is that it's *both* stuff about your muscles *and* stuff about what the rock weighs that explains why you can't lift it. It's *because* the rock weighs what it does that you can't lift it with the muscles you've got. Well, likewise: if you continue to have trouble with S even after we've told you its syntax and lexical inventory, that's surely because there are psychological limits that make it hard to appreciate how the syntax and lexical inventory combine to determine its meaning. To put it another way, if one continues to have trouble understanding S, that's because *an inference is required* to get from a grasp of the lexical/syntactic facts on which its meaning *supervenes* to understanding what its meaning *is.* This is tantamount to endorsing Horwich's 'objection 8', which goes as follows: "The deflationary account fails to do justice to the intuition that we *figure out* the meanings of complex expressions on the basis of our *knowledge* of what their parts mean... [it's] not just that the *facts* about the meanings of primitives determine the *facts* about the meanings of the complexes. It's rather that our *knowledge* of the basic facts must lead by some inferential process to our *knowledge* of what the complexes mean" (171; italics are in the original).

To which objection Horwich replies: "this is indeed a tempting intuition; but [the thesis that understanding complex expression requires inferences] cannot be correct, and so the deflationary attitude should not be faulted for failing to respect it" (171). We'll come in a moment to *why* Horwich says this tempting intuition "cannot be correct". For the moment, we want to go on a bit about just how tempting the intuition is. An analogy should help. Consider claims (i) and (ii) about checking accounts:

i. If you know what your balance was when you started, and what you have deposited, and what you have drawn out, then you know what the balance of your account is.
ii. If you know the things (i) enumerates, you needn't do anything more (in particular, you needn't do anything inferential) to figure out what your balance is.

We take it that (i) is approximately truistic; it follows from *what sort of thing a balance is* (or, if you prefer, it follows from what 'balance' means).[5] Our point, anyhow, is that (ii) doesn't follow from (i) and, moreover, that (ii) is implausible on the face of it.

Why (ii) doesn't follow from (i).

The following schema has the form of an intentional fallacy:

iii. That it is the case that P determines (nomically, metaphysically, or conceptually) that it is the case that Q.
iv. Jones knows that it's the case that P.
v. Therefore: Jones knows that it's the case that Q.

[5]We don't like the parenthesized way of talking; we like to keep our metaphysics clear of our semantics. But the present issues don't in any way turn on that, so we're prepared to be concessive.

Accordingly, the following substitution instance of the schema is invalid:

vi. That one started with three dollars in the account, deposited two dollars and withdrew one dollar determines that the current balance is four dollars.
vii. Jones knows that he started with three dollars in the account... etc.
viii. Therefore: Jones knows that his current balance is four dollars.

We suppose that (ix)-(xi) is likewise invalid and for the same reasons.

ix. That 'John' means *John*, 'loves' means *loves* and 'Mary' means *Mary* (together with syntax) determines that 'John loves Mary' means *John loves Mary.*
x. Bill knows that 'John' means *John*, and that 'Mary' means *Mary,* etc.
xi. Therefore: Bill knows that 'John loves Mary' means *John loves Mary.*

We can now see just why, though there is 'a sense in which' it's sufficient for understanding a complex expression that one grasps its syntax and the meaning of its constituents, there is also 'a sense in which' it isn't. What usually happens when P is metaphysically sufficient for Q is that the inference *believes (…P…)* → *believes (…Q…)* is valid on one way of reading 'believes' but invalid on another.

Let's see where things stand. We're pretty sure that Horwich would agree with the intuition that there's a robust reading of 'know one's balance' on which the inference (vi)-(viii) is fallacious. But we take it that he denies the putative analogy to (ix)-(xi). Why? Well because, in the linguistic case, "transitions between states of understanding do not work in this way, because the beliefs involved in knowledge of meanings are *implicit* [sic]... But since [those beliefs are] implicit, *[and thus consist] in no more than the fact that the expression means a certain thing to him,* its explanation should not be expected to involve inferential processes" (171-172, our emphasis).

So the difference between the checkbook case, where we take it that Horwich accepts the "tempting intuition" that there is figuring out going on, and the language understanding case, where he rejects it, is that whereas the beliefs germane to checkbook balancing are explicit, the beliefs germane to sentence understanding are not. This difference matters, according to Horwich, because of a certain metaphysical truth about tacit knowledge: *Qua ex*plicit, your current belief that you have four dollars in the bank is constituted by your being in a mental state with certain causal powers; presumably the sorts of causal powers that affect "transitions between states of understanding" and that are manifested when you think, talk, etc. about what you have in the bank. But, *qua implicit*,[6] your tacit belief that 'John runs' means *John runs* is constituted simply by the fact that you take 'John' to mean *John* and 'runs' to mean *runs* (and the syntax to be what it is). So, on this account, there's a deep difference between the metaphysics of tacit belief and the metaphysics of explicit belief.

[6]Or perhaps it's *qua* implicit knowledge of one's own idiolect (see Horwich's fn. 14, p.172). We're not clear whether Horwich holds a deflationary view of implicit knowledge *per se*, or just about implicit knowledge of language. Nothing, however, turns on this in the arguments that follow.

Now, to be perfectly frank, we find this all very dark. It is, in particular, quite unclear to us why implicit beliefs should be supposed to differ, in any such way, from explicit ones. For all we know, and, certainly, for all that Horwich has argued, *implicitly* believing that *S means P* and *explicitly* believing that *S means P* might both turn out to be having your brain in a certain functional or neurological state; or they might both turn out to be having a Mentalese sentence that means that *S means P* tokened in your belief box…; or whatever. If any such story is right, then it's not clear why the two kinds of beliefs mightn't be acquired by much the same kinds of inferential processes.

We can now say more clearly what our argument with Horwich is about. We think he is right that there is a kind of case (quite different from checkbook balancing) in which all that's required for grasping something complex is having the right beliefs about its structure and constituents. We take it that Horwich agrees there is a kind of case (of which checkbook balancing is an example) where having the right beliefs about its structure and constituents is *not* sufficient for grasping something complex. However, Horwich thinks the difference between the two kinds of cases is that, in the first but not the second, the beliefs about the constituents of the expression are *implict.* By contrast, we think the difference is that, in the second but not the first, the understanding of the complex is (merely) *dispositional* (where the contradictory of (merely) *dispositional* is something like *occurrent.*)

Notice that, though both apply (*inter alia*) to mental states, implicit/explicit is a quite different kind of distinction from occurrent/dispositional. The former is epistemological; it's a matter of whether the creature that's in a state has (non-inferential) access to its being there. In the simplest examples, implicit/explicit is about whether a creature is able to report being in the state that it's in. Whereas the second distinction is *ontological; w*e suppose that *occurrents*, but not dispositions, are species of events. That is part and parcel of the fact that they are associated not just with *stretches* of time, but also with *instants*. If John's thought that the cat is trapped in the closet is occurrent, then 'when did it occur to him?' presumably has an answer ('at 3:17'; 'when he first heard the cat say meow' and so forth.) But if his thought that the cat is trapped in the closet is merely dispositional (as in the case where John is congenitally disposed to occurrent *cat-in-the-closet* thoughts) the pertinent question is not 'when did he have it?' but 'how long did it last?' ('How lonk haf you been vorryink about die gats' beink in de gloset, Mister Portnoy?) We're aware, of course, that it's in dispute just how the distinction between dispositions and occurrents should be drawn, and we don't want to get involved in the argument. Suffice it that it's metaphysical rather than epistemic by general consensus; hence, by general consensus, different from implicit/explicit.

So, then, our story is that one's understanding of a sentence can be any combination of explicit/implicit with occurrent/dispositional (except that an explicit mental state presumably has to be occurrent; see fn. 6) In the case where one's understanding of an expression is implicit and (merely) dispositional, Horwich may well be right that it comes to no more than one's grasp of the syntax and the meanings of the parts. However, we think that's because such cases are *dispositional,* not because they're *implicit*. That is, like all

our cognitive scientist friends, we think there is such a thing as understanding that is implicit but *occurrent* (a species of unconscious mental process, we suppose). Certainly Horwich hasn't given any reason to doubt that there's such a thing. Nor has he given any reason to believe that, when implicitly understanding a sentence is an occurrent process, inferring the sentence's compositional structure is other than essential.

We're pretty sure we have this stick by the right end since we can't think of any reason why an occurrent belief shouldn't be arrived at inferentially, whether or not it's explicit. By contrast a (merely) dispositional belief can't be arrived at *inferentially* because it can't be *arrived at* at all. There might, of course, be mental processes that cause you to have a (merely) dispositional belief; and some of the episodes in such mental processes might consist of having thoughts occur to you. But these thoughts wouldn't count as premises from which the dispositional belief is *inferred*. You can't *infer* that P unless it (implicitly or explicitly) *occurs to you* that P; and (merely) dispositional beliefs are *ipso facto not* ones that (implicitly or explicitly) occur to you.

Here's where we've got to now. It's urged against the deflationary account of compositionality that it "fails to do justice to the intuition that we *figure out* the meanings of complex expressions on the basis of our *knowledge* of what their parts mean..." Not so, Horwich replies; though it's tempting to think that understanding complex expressions depends on inferences from their syntactic and lexical constituency, that thought must be resisted when the beliefs involved are implicit. Horwich offers no argument for this, however, and he needs one badly. For, the obvious candidates for non-inferential beliefs about the meanings of complex expressions aren't the *implicit* ones, they're the (merely) *dispositional* ones. If that's right, then there is no reason so far why some of one's beliefs about expressions shouldn't be both implicit and occurrent. And there is no reason so far why these *implicit*, occurrent beliefs about expressions shouldn't be much like one's *explicit*, occurrent beliefs about one's balance; *viz.,* both inferential. We suspect, in short, that Horwich has confused the distinction between implicit and explicit beliefs with the distinction between (merely) dispositional and occurrent beliefs, and that this has lead him to argue, in effect, that since when beliefs about expressions are implicit they can't be inferential. This argument is ok if you replace 'implicit' by 'merely dispositional', but there's no reason to believe it as it stands.[7]

But, we're still not out of the woods. For, according to Horwich there is an independent argument against the "tempting intuition" that sentence understanding, like checkbook balancing, can involve a lot of 'figuring out', i.e., an argument that *doesn't* assume that implicit knowledge is *ipso facto* non-inferential, but that shows all the same that "...at the fundamental level, compositionality is not explained in terms of inference" (174). Here's the argument: "...although there may be *some* language whose complex expressions are understood as a product of explicit inference, such inferences would have to take place in a more basic language whose complexes would themselves already have to have some

[7]Notice that whereas (according to us) implicit beliefs may nonetheless be occurrent, it's presumably a truism that explicit beliefs can't be merely dispositional. Perhaps it's because the two distinctions fail, in this respect, to be independent that Horwich got confused between them.

content or meaning; and if inferences are required for this, then a yet more basic language would be needed in which to conduct them... and so on" (172).[8]

This regress argument has been around for a long time; probably it was invented by someone who lived in a cave. The usual reply strikes us as convincing: by assumption, understanding English expressions is inferential because you have to translate them into some other language (into Mentalese as it might be) in order to use the information they convey. But you don't have to translate Mentalese in order to use the information it conveys. All you have to do is think in it. So there isn't a regress after all.

But Horwich isn't having that. "It is all very well to refuse to speak of 'understanding' and 'possession' of meaning in connection with the language of thought, and thereby to hope to retain the idea that when a complex is, properly speaking, 'understood', inference is invariably involved...[But that] merely obscures the fact that the same issues [about compositionality] arise with respect to Mentalese, but in a slightly different formulation" (173).

Let's, please, be very careful about what's at issue here. To begin with, we're quite prepared to concede, for purposes of the argument, that someone who thinks in Mentalese thereby counts as understanding it 'in some sense'. *Pace* Horwich, what we've called "the usual reply" to the regress argument is not supposed to be terminological. In particular, it's not supposed to turn on whether what one does with Mentalese counts, *strictu dictu*, as understanding it. Rather, the issue is whether assuming that English and Mentalese are both compositional, and that understanding English is inferential, requires assuming that whatever constitutes understanding Mentalese is inferential too. Horwich apparently thinks we must; we think we needn't, and we affirm that we don't.

The following view seems to us coherent and neither vacuous nor gratuitous: English is compositional, and the process of understanding its sentences is inferential. But the fact that the process of understanding English sentences is *inferential doesn't follow from the fact that English is compositional.* So far, then, there's nothing to prevent a language from being compositional even though the process of understanding its sentences *isn't* inferential. If this possibility is coherent and begs no questions, then assuming that understanding English is inferential, *and* that all inference requires a linguistic vehicle, *and* that one thinks in sentences of Mentalese implies no regress so far.

Can one conjure up a plausible story according to which all of that is true? Sure; in fact, it's just the standard language of thought story. Readers who have already heard this story, and are tired of it, are advised to skip directly to Part 2.

We need, in particular, two assumptions:

[8]There's something puzzling about this formulation. Presumably, it's a mistake for Horwich to suggest that the inferences that generate the regress have to be *explicit*. If a regress threatens, it's because of the putative necessary connection between understanding a language and making *any inferences at all*, explicit or otherwise. We'll take this reading for granted in what follows.

xii. The mental processes that are the consequences of understanding a sentence are mediated by a mental representation which displays its logical form; this is true both of English and of Mentalese.
xiii. Sentences of Mentalese are explicit about their logical form in ways which, notoriously, those of English are not).

If (xii) and (xiii) are both true, then inference might enter into one's use of English in a way that it doesn't enter into one's use of Mentalese, even though both Mentalese and English are both compositional. That's because, in the case of Mentalese but not in the case of English, the truth of (xiii) assures that (xii) is satisfied vacuously. The trick is to motivate the claim that understanding English is inferential *independently* of motivating the claim that it is compositional, thus leaving open the possibility that Mentalese might be compositional even though inference isn't required to use it. As far as we can tell, Horwich offers no argument at all against this tactic; it appears, indeed, that he is unaware of its existence as a polemical possibility.

Notice that we don't have to show that there is such a language in order to undermine Horwich's regress argument; all that's required is that (xii) and (xiii) are coherent. Still, as a matter of fact, our main reason for thinking they are coherent is that we think they are probably both true. So we'll say a little about that.

-*Why you might want to endorse what (xii) says about English.*

Because you think that understanding an English sentence involves representing it in a way that formally determines its entailments, and that logical form does so but surface structure doesn't.

Because you think that understanding an English sentence requires (*inter alia*) recovering an ambiguity-free representation of the sentence; *afortiori*, it requires recovering a representation which distinguishes ambiguities of logical syntax.

Because you hold a computational view of mental processes according to which the consequences of understanding an English sentence are determined by mental operations that are sensitive to logical form of the sentence.

There are other reasons too. But this should do to be getting on with.

-*Why you might want to endorse what (xii) says about Mentalese.*

Because you hold a view of mental processes according to which the psychological consequences of tokening a Mentalese sentence are causally determined by operations that (a) are responsive to its logical form and (b) apply to the sentence in virtue of its syntax. Making the logical form of a Mentalese sentence explicit in its syntax is how to meet both these conditions at once. As, indeed, Turing taught us.

-*Why you might want to endorse (xiii).*

Precisely in order to avoid the regress that threatens if you have to *infer* the logical syntax of a Mentalese sentence in order for it to play its characteristic causal role. Roughly, if (xiii) is true, then all that is required for it to play this role is that it be tokened (e.g., in the belief box.) That's *why* talk of understanding a Mentalese sentence is otiose in a way that talk of understanding an English sentence is not.

So then: We hold that no regress threatens the view that English and Mentalese are both compositional and that understanding English requires a kind of inference and that thinking in Mentalese does not.

Suppose, however, that we're wrong about all this and there is, after all, some way to show that language understanding isn't inferential. What would that imply about the deflation of compositionality? In particular, what would it imply about whether the compositionality of a language constrains the semantics of its lexicon? We think the right answer is 'Nothing.'

The patient reader will remember that Horwich has two formulations of his "main thesis", these being that "compositionality of meaning imposes no constraint at all on how the meaning properties of words are constituted" and that "understanding a complex expression (non-idiomatically) is, by definition, nothing over and above understanding its parts and knowing how they are combined." We now return to a point that we made at the outset: Though Horwich asserts them interchangeably, these two formulations don't appear to be equivalent; in particular, the second doesn't appear to imply the first. Correspondingly, if it's the first that one really cares about, it doesn't matter whether the second is true. So then, after all this ground clearing, we propose to put the issues about how (/whether) understanding sentences requires making inferences entirely to one side. That allows us to turn directly to the question: 'does an account of the compositionality of a language constrain the nature of its lexicon?' We'll now argue that *of course it does.*

Part II: Compositionality and the Lexicon

We started this paper by rehearsing a familiar informal construal of the notion of compositionality: in effect, that the meanings of complex expressions supervene on their syntax together with the meanings of the lexical primitives they contain. We remarked that if a language is compositional in that sense, then, *prima facie*, that imposes some quite significant constraints on what the meanings of its primitives could be. It's hard to see, for example, how primitive meanings could be birds or chairs since, whatever complex meanings may be it's hard to see how birds or chairs could be parts of them.

You might suppose that Horwich would find this line of thought anathema; but he needn't, and as far as we can tell he doesn't. For, our way of putting the point assumes that the meanings of complexes expressions contain the meanings of primitive expressions (so that the meaning of 'loves' is part of the meaning of 'John loves Mary' and so on.) Such assumptions are not, however, entailed by the supervenience thesis *per se*. Supervenience *per se* entails only that *whatever the semantic facts about complex*

expressions may turn out to be, they are determined by their syntax together with the semantic facts about their lexical primitives, whatever *they* may turn out to be.

Horwich's deflationary story about compositionality is only this: compositionality places no constraint on primitive meanings *if one prescinds* from all assumptions about *complex* meanings except that they supervene on syntax and lexical inventory. For, as Horwich says, "...*whatever* their underlying nature may turn out to be, there are bound to be construction properties (of the form x [= the meaning of the complex] results from applying procedure p to primitives whose meanings are, $<W_1,...,W_n>$). Hence ... it is bound to be the case that the facts regarding the meanings of the complex expressions are derived from facts about the meanings of the primitives" (160). We take this to be just the point that, *if it's left open* what the semantic facts about the complexes are, then, whatever the meanings of the primitives are, there is sure to be some way of mapping the latter onto the former. If *that's* all that semantic compositionality requires, then, as Horwich says, it can't but be true.

But all that shows is that the unexiguous notion of compositionality that follows from supervenience alone isn't the robust notion of compositionality that a theory of sentence meaning requires. And, if that's all that the deflation of compositionality amounts to, it's of no great interest that compositionality deflates. People who think that compositionality substantively constrains lexical meaning have it in mind that there are all sorts of presumptive truths about the semantics of complex expressions that need explaining; and that it's precisely the assumption of compositionality, together with a theory of the primitive meanings, that is supposed to explain them. Such presumptive truths about complex meanings as these, for example:

-Complex meanings are semantically evaluable (e.g., for truth or satisfaction).

-Although the syntax and lexicon of English are finitely specifiable, there is a denumerable infinity of distinct complex meanings.

-There are n-complex meanings for each intuitively n-ways ambiguous complex English expression.

-One meaning of the sentence 'John ate his peas' is such that 'John' has scope over 'his'.

-The meaning of 'John snores' and the meaning of 'John swims' are such that both sentences make reference to John.

And so on and on. And on.

Consider, for a further example, the arguments about systematicity that are currently live in cognitive science. Roughly, systematicity is the fact that any language (/mind) that can express (/entertain) the proposition P will also be able to express (/entertain) many propositions that are semantically close to P: Anyone who can think the thoughts that *John snores* and that *flounders swim* can likewise think the thoughts that *flounders snore*

and that *John swims*. (Likewise, *mutatis mutandis* for understanding sentences of a language that can express these thoughts.)

It's pretty widely agreed that an explanation of the fact that complex meanings are systematic requires assuming that lexical meanings are context independent. The idea is this: compositionality says that the meaning of 'John snores' and of 'John swims' depend, *inter alia*, on the meaning of 'John'. And it's because 'John' means the same in the context '...snores' as it does in the context '...swims' that if you know what 'John' means in one context you thereby know what it means in the other.

So compositionality, together with the systematicity of complex meanings, places a context-independence constraint on the properties of lexical meanings. This constraint is *highly substantive*. For example, it rules out the theory, held practically universally in the cognitive psychology community, that concepts are stereotypes. The argument goes like this: the systematicity of complex meanings requires the context-independence of lexical meanings; stereotypes aren't context independent (for example, the stereotype of people-swimming is much different from the stereotype of flounder-swimming since the latter, but not the former, adverts to the exercise of fins); so lexical meanings can't be stereotypes.

This seems a pretty good example of how compositionality, together with other considerations about complex expressions, constrains the semantics of primitive expressions. Horwich, considering this case, replies that the argument from compositionality to concepts, lexical meanings, etc., not being stereotypes presupposes a 'uniformity thesis'; *viz.,* that if the meanings of the primitives are stereotypes (or uses, or prototypes, or inferential roles, or whatever), then the meanings of the complexes are *also* stereotypes (uses, prototypes, inferential roles, etc.). Well, it doesn't since, as we've just seen, a context-independence thesis would do equally well; and context-independence is a property that compositionality imposes on the lexicon *whether or not* uniformity is assumed.

Anyhow, there is an independent argument for the uniformity principle. Compositionality says, roughly, that its syntax and its lexical constituents determine the meaning of a complex expression; it's thus part of the explanation of why practically everybody who understands 'dogs' and 'bark' understands 'dogs bark.' But it also needs explaining that you practically never find people who understand 'dogs bark' but don't understand 'dogs' or 'bark'. What we'll call '*reverse'* compositionality explains this by assuming that the meanings of constituent expressions supervene on the meanings of their complex hosts. If that's right, then if you understand 'dogs bark,' it follows that you know everything to determine the meanings of 'dog' and 'bark'. Fine so far, but now there's a further puzzle: as far as anybody knows, compositionality and reverse compositionality *always go together.* Just as you won't find a language which can talk about dogs and barking but can't talk about dogs barking, so you won't find a language which can talk about dogs barking but can't say anything else about barking or about dogs. It would be nice to have an explanation of why the meanings of complex expressions supervene on the meanings of their parts; and of why the meanings of parts supervene on the meanings of their complex hosts. And it would be still nicer if the explanation of these two superveniences

also explained why they always turn up together. In fact, the explanation is obvious; the meaning of 'dogs bark' supervenes on the meanings of 'dogs' and 'bark' because the meanings of 'dogs' and 'bark' are parts of the meaning of 'dogs bark'; and *meaning of 'dogs' and 'bark' supervene on the meaning of 'dogs bark' for exactly the same reason.* But the idea that complex meanings (don't just supervene on, but actually contain) constituent meanings, is the 'uniformity thesis' in a very strong form. So it looks like the uniformity thesis must be true. So it looks like compositionality (together with reverse compositionality, together with the lack of an alternative explanation) severely constrains the lexicon after all; for example, it entails that lexical meanings can't be stereotypes.

This sort of argument ramifies in interesting ways. The meanings of 'dogs' and 'bark' must be contained in the meaning of 'dogs bark' because people who understand the sentence likewise understand the words. But the meaning of 'dogs bark' must be contained in the meaning of 'dogs bark and cats purr' because people who understand the conjunctive sentence generally understand both conjuncts.[9] In fact, the reverse compositionality of complex expressions relative to their *lexical* constituents, is just a special case of the reverse compositionaliy of complex expressions with respect to their constituents *tout court*, lexical or otherwise. Since in natural languages, every constituent expression has infinitely many hosts, this amounts to an infinite amount of reverse compositionality, all of which is, as far as anybody knows, inexplicable unless the 'uniformity condition' is assumed. (For further discussion of the implications of reverse compositionality see, Fodor, 1998, chs. 4,5; Fodor (forthcoming).)

The point we want to emphasize, however, is not that the reverse compositionality argument against stereotypes as lexical meanings is correct (though it is). Our point is that people who think it matters to the lexicon whether complex meanings are compositional have it in mind to deploy arguments whose premises also include many other premises about the semantics of such expressions; that sentence meanings are systematic, that languages and conceptual systems are reverse compositional, that complex meanings are uniform with lexical meanings, etc. It's quite true, as Horwich says, that compositionality doesn't matter much lacking such assumptions. But it's also true that it doesn't matter much that compositionality doesn't matter much lacking such assumptions. What matters is that there appears to be a plethora of truths about the semantics of complex expressions that the assumption of compositionality, together with a good theory of the lexicon, explains; and that, as far as anybody knows, can't be explained if the assumption of compositionality is left out

Summary and Conclusion

This was the burden of Part 1: The standard reasons for holding that understanding English sentences requires making inferences are all basically 'poverty of the stimulus' arguments. They depend on assuming, on one hand, that the surface structure of English sentences is generally inexplicit about semantically salient properties like logical form; and, on the other hand, that properties of the logical forms of sentences, play are essential

[9]It's not, of course, a *necessary* truth that if you understand a syntactically conjunctive sentence you understand each syntactic conjunct; the sentence might be an idiom.

determinants of their causal consequences of their tokenings. That is all quite compatible with supposing that, although Mentalese is compositional, you don't have to understand it in order to think in it.

This was the burden of Part 2: The standard arguments that run from compositionality to the nature of lexical meaning turn on the need to explain such familiar properties of complex meanings as productivity, systematicity reverse compositionality and the like. It is therefore not surprising, and not awfully interesting, that compositionality is deflatable when one prescinds from its role in such explanations. Why, after all, should anyone *want* to prescind from the role of compositionality in explaining systematicity, productivity and the like? Natural languages, and human minds, *are* systematic and productive, and that they are needs explaining.

Anyhow, as far as we can tell, the argument discussed in Part 1 that understanding English requires making inferences, and the argument discussed in Part 2 that compositionality constrains lexical semantics, are independent in both directions. Thus, the compositionality argument that shows that lexical meanings can't be stereotypes applies both to English and Mentalese, even though, by assumption, using the former requires making inferences but using the latter doesn't. Conversely, there presumably could be language whose use has to be inferential (e.g., because its formulas aren't explicit about their logical forms) but which is none the less not compositional. A finite language, all of whose expressions are idioms, would do the trick.[10] The long and short is: All that the claim that understanding English requires inference and the claim that compositionality constrains lexical semantics have in common is that there are convincing arguments for each.

[10]Because he holds that compositionality places no substantive constraints theories of language, Horwich is presumably required to hold that being compositional is not a property that languages have *contingently.* (See his discussion of objection 5.) It would seem to follow that the meanings of complex expressions must supervene on their syntax and lexical contents *even in a finite language.* This strikes us as a *reductio*; surely a finite language could consist only of idioms?