

## Attributions of Meaning and Content

1. Consider schemas like

(M<sub>S</sub>)  $\lambda p=$  means that  $p$ ,

where  $\lambda p=$  is to be replaced by any meaningful sentence of our language; or more generally,

(M)  $\lambda e=$  means  $\langle e \rangle$ ,

where  $\lambda e=$  is to be replaced by any meaningful expression of our language. The bracket quotes are an invented notation, but

$\lambda \text{Hund}=$  means  $\langle \text{dog} \rangle$

is intended to be simply a formal representation of our usual devices of meaning-attribution for subsentential expressions; it is intended to be neutral to the question of what sort of entity if any the bracket-term refers to. In the case of sentences we don't need special bracket terms, because we have  $\lambda \text{that}=\text{clauses}$ , but we can regard phrases of form  $\langle p \rangle$  as simply an alternative notation for the corresponding phrases of form  $\lambda \text{that } p=$ , so that (M<sub>S</sub>) is simply a special case of (M).

I should say something about the use of the schematic letters  $\lambda p=$  and  $\lambda e=$  in the schemas above. My focus will be on the individual thinker. An individual thinker has an idiolect—a set of expressions that he or she understands (in a way that may or may not accord with how others understand it) which will typically not coincide with the boundaries of any language, or of the union of any languages. I will take the appropriate substitution instances for the schematic letters in the above schemas for a given person (at a given time) to be the expressions *in that person's idiolect* (at that time) that are of the appropriate syntactic category. (Understanding comes in degrees, so the boundaries of a person's idiolect are somewhat vague, and so there is some vagueness as to which substitution instances are appropriate. I don't think that this fact will matter much in what follows.)

Of course, if there are ambiguous expressions in the person's idiolect, the above schemas need to be adjusted for this. One way to do this would be to rewrite them as

$\lambda p=$  (as I'm now understanding it) means that  $p$

and

$\lambda e=$  (as I'm now understanding it) means  $\langle e \rangle$ .

Some would want to explain ways of understanding sentences in terms of propositions (and uses of subsentential expressions in terms of parts of propositions—concepts, in one sense of that term). I don't refuse to talk in terms of propositions in all contexts, but I think that such talk often obfuscates the

issues, and that it would do so here. When a sentence in a speaker's idiolect is ambiguous, the speaker processes different tokens of it in systematically different ways: for instance, some tokens of >Visiting relatives can be boring= have a direct inferential connection to >My cousin is usually boring when he visits me= and others don't. We don't need to invoke relations to distinct propositions to make the distinction between the two readings of this or any other ambiguous sentence. (The point can be extended to sentences that contain indexicals: two of my tokens of >She is beautiful= said of obviously different women have different readings in that they are linked up with distinct inner file drawers. I'm using >means= in such a way that they differ in what they mean.)

Instead of invoking *ways of understanding*, it might be clearer to stick to the original versions of (M<sub>S</sub>) and (M), but stipulate that the quotation mark name refers not to an *orthographic* type but to a *computational* type: it refers to a class of expression tokens (or potential tokens) in the meaning-ascriber's actual idiolect that are treated computationally as equivalent. This is somewhat reminiscent of the dot quotes employed in Sellars 1962. But those were supposed to refer to *interpersonally ascribable* conceptual roles, and thus give rise to Quinean problems about which beliefs affect conceptual role and which don't. The notion of sameness of computational role I'm employing should be regarded as meaningful only for expressions in the same idiolect.

Clearly the schema (M) and its special case (M<sub>S</sub>) have a special epistemological status: they are not open to serious doubt on empirical grounds. *Prima facie*, this is somewhat puzzling: it certainly isn't beyond doubt or empirical discovery what a word in *another* language means; how then can it be beyond doubt or empirical discovery what a word in *my* language means? (Of course, there is an obvious sense in which what it means in my language isn't beyond empirical discovery: I might learn that >metacompact= doesn't mean quite what I suppose it does. But the point is that I can't learn that it doesn't mean <metacompact>.)

What explains the special epistemological status of the schema?

2. One possible answer is *the linguistic view of meaning attributions*. Putting aside doubts about the clarity of interpersonal synonymy for now, here's an initial way to put the view: to say that a word means <rabbit> is simply to say that it means the same as >rabbit= as I actually understand that word. (Or as I'm actually understanding it *now*, if the word is ambiguous.) The >actually= is there to handle counterfactual contexts: in a world like the actual world except that the English usage of >rabbit= and >cockroach= are interchanged, >cockroach= and the French >lapin= mean the same as >rabbit=

*actually* means for me, so they mean <rabbit>.

The key idea is that although a meaning attribution may appear to relate a word to a mind-independent entity (a meaning), it really just relates the word to one of the meaning-attributor's words (computationally typed). Or perhaps, to one of the meaning-attributor's mental symbols that is causally correlated to his use of the word; I'll count that as not essentially different.

Of course, the relation of the foreign word to my word (or my mental symbol) is based on characteristics of each word separately, e.g. on the inferences that govern certain kinds of sentences containing the words and on the worldly conditions that typically lead to the assent to other kinds of sentences involving the words. Let's call such characteristics of a word (on which translation of the word or synonymy judgements about it are based) its *meaning-characteristics*. To make this explicit, we can restate the linguistic view roughly as follows: *to say that a word means <rabbit> is just to say that it has the same meaning characteristics as the meaning characteristics of my actual use of »rabbit«*. Here the meaning-characteristics of a word do not include relations to intentional entities (meanings or concepts); indeed, I take it to be possible to describe the meaning-characteristics of a word without using any semantic terms at all, including terms like »refers«.

This may need a small refinement: the requirement of *exact* sameness of meaning characteristics is obviously appropriate only if you understand »meaning characteristics« in a suitably coarse-grained sense. So as not to require this special coarse-grained notion, we can restate the idea in terms of a special equivalence relation *M* on the set of possible meaning-characteristics, finely described:

(LV) To say that a word means <rabbit> is just to say that it has meaning characteristics that are *M*-equivalent to the actual meaning characteristics of my term »rabbit«.

Of course, interpersonal sameness of meaning is at least somewhat indeterminate, which means that if we do things this way, we should recognize that it is indeterminate precisely which equivalence relation »*M*-equivalence« should be taken to stand for. Indeed, on a more Quinean view the problem with interpersonal synonymy goes beyond mere indeterminacy; on such a view, we should avoid mentioning equivalence relations at all, and just say something like:

(QLV) To say that a word means <rabbit> is to say that its meaning characteristics make it appropriate to translate the word into my actual language as »rabbit«.

I will not decide between these two variants of the linguistic view.

If the linguistic view (in either form) is acceptable, then it is obvious why the meaning schema

should have special epistemic status: its instances in effect have the form

(\*) p if and only if actually p,

whose special epistemic status is generally recognized and understood. (More accurately, they have the form of generalizations of sentences of form (\*): they say that for any meaning characteristic C, my word has C if and only if it actually has C.)

It seems clear that any considerations favoring the linguistic view of meaning attributions would equally favor a linguistic view of saying-that attributions, belief-attributions, desire-attributions, knowledge-attributions, and so forth.

3. There are a number of possible objections to the linguistic view of meaning-attributions.

One is the Church translation argument (Church 1950): it is alleged that if A>lapin= means <rabbit>@says that >lapin= means the same as >rabbit=, then its German translation ought to be A>lapin= bedeutet <rabbit>@rather than A>lapin= bedeutet <Kaninchen>@, which goes against both the practices and purposes of normal translation. (Recall that the brackets are just my notation for ordinary meaning-attributions; if sentences were in question, I could just use >that= clauses.) Schiffer (1987) poses the point slightly more sharply: simplifying his formulation a bit, let's imagine that Karl and Fritz are monolingual German speakers, and Karl is told that Pierre=s utterance means the same as >Der Schnee ist weiss= while Fritz is told that it means the same as >Snow is white=. Schiffer argues in effect that if we were to interpret AS said that snow is white@linguistically, as AS said something equivalent to >Snow is white=@, this would have the absurd implication that it is Fritz rather than Karl who believes that Pierre said that snow is white.

In my view, the Church-Schiffer argument against the linguistic view shows only that the view needs to be formulated a bit more carefully. Before saying what the argument does and doesn't show, let me tell a little story. Imagine an English novelist whose books are translated into French. He is dismayed to discover that in the French translation of the first novel, the translator (a philosopher of language) left the quotations untranslated, on the ground that when the novel said that the character uttered a certain English sentence it would be a mistranslation to have the character uttering a French sentence. To guard against such bloody-mindedness, the novelist in his second novel used special quotation marks, #-quotes, when he wanted the dialog translated. (He decided that the need to resort to this device had unexpected benefits, because occasionally he wanted English quotations left untranslated, and this allowed the translator to know which was which.) Of course, he had underestimated just how bloody-minded philosophers of language can be: his translator still left the

quotations untranslated, saying that whatever the author wanted, the #-quotes still referred to English sentences and so it would be a mistranslation to translate them otherwise. So for his third novel, the author told his publisher not to hire a translator, hire instead a quasi-translator who will translate within #-quotes.

The advocate of the linguistic view as described above is like the novelist: he is undone by bloody-minded philosophers of language. The idea behind the linguistic view was that the referent of an English >that=-clause is the sentence contained in it, and more generally the referent of a bracket-quotation is the same as an ordinary quotation; semantic claims like

(1) >Lapin= means <rabbit>

and

(2) >La neige et blanc= means that snow is white,

which appear to relate French expressions to intentional entities, really just relate them to English expressions. (Similarly, the sententialist will hold that when we say that a person says that snow is white or believes that snow is white, we are expressing a relation between the person and the sentence >Snow is white=.) So stated, I think the idea is quite defensible; but what needs to be added is that the bracket-quotes and >that=-constructions *don't* behave quite like ordinary quotation marks (at least as those are standardly construed by philosophers of language), they work instead like the novelist's #-quotes. This means of course that we want them to be quasi-translated rather than *Aliterally translated*@ (on the Church-Schiffer understanding of what literal translation involves). But this point about the kinds of translation we want should be taken as reflecting a view of what features of language we care about. We *don't* care much about *Aliteral meaning*@, if that is what is preserved in *Aliteral translation*@ as conceived by Church and Schiffer; what we care about, rather, is what is preserved in quasi-translation. We might in deference to Church and Schiffer call this quasi-meaning, though I think it is what most would simply call meaning. The rationale for our interest in *Aquasi-meaning*@ rather than (what I reluctantly call) *Aliteral meaning*@ is that what each of us must do in understanding foreign utterances is to compare them with utterances we already understand, and such comparisons are what is preserved in quasi-translation but not in *Aliteral translation*@.

To summarize, one key part of sententialism is that our interest is in *Aquasi-translation*@ and *Aquasi-meaning*@ rather than *Aliteral translation*@ and *Aliteral meaning*@. The other key part is that the way to understand (2) is as

(2=) >La neige et blanc= means the same as #Snow is white#;

or better, as

(2 $\Rightarrow$ )  $\lambda$ La neige et blanc=quasi-means the same as  $\#$ Snow is white $\#$ ,

where again, quasi-meaning is what is preserved in quasi-translation. The use of #-quotes rather than ordinary quotes around  $\lambda$ Snow is white= guarantees that the quasi-translation of (2 $\Rightarrow$ ) or (2 $\Rightarrow$ ) into German involves  $\lambda$ Der Schnee ist weiss $\#$ . The use of  $\lambda$ quasi-means= instead of  $\lambda$ literally means= in (2 $\Rightarrow$ ) is unimportant when discussing  $\lambda$ La neige et blanc=, since  $\lambda$ literal meaning@ and quasi-meaning coincide for that sentence, but it becomes important when a sentence with #-quotes is under discussion. (This includes a sentence with a  $\lambda$ that= clause or bracket-quotes, given the linguistic construal of these.) The fact that it is quasi-meaning rather than meaning that is important to the sententialist is one good reason for preferring (2 $\Rightarrow$ ) to (2 $\Rightarrow$ ) (or rather, to (2 $\Rightarrow$ ) with  $\lambda$ means= interpreted as  $\lambda$ literally means@); another good reason is implicit in the next paragraph.

I claim that this version of the sententialist doctrine is immune from the arguments of Church and Schiffer.<sup>1</sup> The point with regard to Church is too obvious to require comment. The Schiffer argument is a bit more interesting, in turning on a double intentional context: believing that a person says that p. Clearly any sententialist who elevates quasi-meaning over meaning will hold that to believe that q is to accept a sentence that quasi-means the same as  $\#$ q $\#$ . So to believe that Pierre said that snow is white is to accept a sentence that quasi-means the same as

(3) Pierre said that snow is white;

on our sententialist doctrine, that will be a sentence that quasi-means the same as

(4) Pierre said something that quasi-means the same as  $\#$ snow is white $\#$ .

Since clearly it is Karl rather than Fritz who accepts such a sentence, the Schiffer argument has no force against this version of the linguistic view.

**4.** A rather more interesting set of issues to raise about the linguistic view is its implications about explanation: in particular, the fact that if we explain what a person does in terms of her belief that p, the linguistic construal has it that we are in a sense bringing ourselves into the explanation: we are saying that the person accepts a sentence (or mental representation) which has a role in her psychology like the role that the sentence  $\lambda$ p= (or a mental representation that we associate with it) plays in ours. I don't think this consequence of the linguistic view is terribly upsetting: in explaining the behavior of things of whose detailed workings we are ignorant, we often bring in other things we assume to work in roughly analogous ways. (AI bet your car is acting up like that because it has the same engine problem that my car had last week.@) In fact, I think it a virtue of the linguistic view that it brings out this feature of

intentional explanation; but I can not discuss these issues here. (See the Postscripts to Chapter 2 and Chapter 4.)

Instead, I will simply note an initial reason for thinking that it is not propositions, but rather sentences we understand, that do the work in explanations of what people do. For in order to use Rebecca's beliefs in any normal explanation of her behavior you must be able to supply a sentence you understand (whether in public language or Mentalese) that expresses her belief: you must be able to say *Ashe believes that p*, where *p* is a sentence you understand. Being able to designate the proposition Rebecca believes in some other way (e.g. via a sentence you don't understand, or a list of possible worlds, or a name like Horace) would be of no use to us in our normal explanatory practice. (Objection: *That's because these wouldn't tell me which proposition it is.* Response: Maybe, but if so, that's because our standard for *knowing which proposition it is that the belief expresses is being able to supply a sentence we understand that expresses the proposition.* That should give the game away: it is the sentence we understand that does the explanatory work.)

5. One might also object that on this view of meaning, to say that *Der Schnee ist weiss* means that snow is white will not imply, without auxiliary premises, that it is true if and only if snow is white. (It will of course imply that it is true if and only if our sentence *Snow is white* is true; but the claim that our own sentence *Snow is white* means that snow is white will be trivial on this view, and won't imply that our own sentence is true if and only if snow is white. So saying that *Der Schnee ist weiss* is true if and only if *Snow is white* is true won't imply, *without further information*, that it is true if and only if snow is white.) Because of this, it might be objected that on this account, attributions of meaning to sentences leave out a key ingredient of meaning, viz. truth conditions. Similarly, it might be objected that on this account, the description of the meaning of a singular term or predicate has no implications for the conditions under which it refers to or is true of a given object, again leaving out a key ingredient of meaning.

I don't really see why an attribution of meaning to *Der Schnee ist weiss* must imply that it is true if and only if snow is white *all by itself*, i.e. without use of the uncontroversial background fact that *Snow is white* is true if and only if snow is white. But even if one accepts that it must, an advocate of the linguistic theory has an easy response if he adopts a disquotational theory of truth, reference and the like. According to this disquotational view, the claim that *Snow is white* is true if and only if snow is white, and the claim that *the teacher of Alexander* refers to something if and only if that person was the teacher of Alexander, are not contingent empirical claims about our language, they are analytic

truths. So when I say that a foreign term is synonymous with 'the teacher of Alexander', I imply *in a quite literal sense* that the foreign term refers to an object iff that object is the teacher of Alexander, and so the objection that the view divorces meaning from reference has no weight.

Indeed, it is worth noting that an advocate of the linguistic view can adopt a version of disquotationalism that is often regarded as more plausible than standard versions. Standard versions of disquotationalism take as central such schemas as

(T) 'p' is true if and only if p

and

(R) For any x, 'b' refers to x if and only if b=x.

But an advocate of the linguistic view can instead take as central such schemas as

(T\*) It is true that p if and only if p

and

(R\*) For any x, '<b>' refers to x if and only if b=x;

for if we adopt a linguistic view of 'that'-clauses and other bracket terms, these amount to the same thing. The only advantage of (T) and (R) over (T\*) and (R\*) is in making unambiguously clear that it is our own sentences and expressions that we are talking about, rather than special intentional entities.

The equivalence of (T) and (T\*) on a linguistic view implies that they have the same modal status. In my favored version, they are both necessary. The view of quotation mark names as naming computational types can be used to motivate this. For just as sameness of computational type seems ill-defined across speakers, it seems ill-defined across possible worlds, or at the very least, across possible worlds where our use of language is importantly different. To say that at another possible world *w* a computationally typed sentence is true can only mean that the sentence *as understood at our world* is true-at-*w*. So to say that ~('Snow is white' is true iff snow is white), i.e. that at all worlds *w*, 'Snow is white' is true iff snow is white, should be understood as meaning that for all worlds *w*, 'Snow is white' as understood at our world is true at *w* iff at *w* snow is white. And this is correct.

The linguistic view also allows us to put the deflationary view in a more general way, applicable to arbitrary languages (and to our own language in other possible worlds): the generalized versions are

(T<sub>G</sub>) For any sentence *S* (of any linguistic community *C*, in any possible world *u*), if *S* means that *p* (for *C* in world *u*) then *S* (as used by *C* in world *u*) is true (at world *w*) if and only if *p* (at *w*).

and

(R<sub>G</sub>) For any term *t* (of any linguistic community *C*, in any possible world *u*), if *t* means  $\langle b \rangle$  (for *C* in world *u*), then for any *x*, *t* (as used by *C* in world *u*) refers to *x* (at world *w*) if and only if *x* is *b* (at *w*).

These together with the appropriate instances of (M) (namely,

$\rangle p =$  means that *p*

and

$\rangle b =$  means  $\langle b \rangle$ )

give rise to (T) and (R). This more general formulation of disquotationalism has advantages; its disadvantage is that the unwary might fail to see that it contains a hidden bias toward our own language, which is here built into the notion of meaning that *p* or meaning  $\langle b \rangle$ . (The generalization is in effect the same as the Quasi-disquotational truth discussed in the previous chapter, except that here I have been neutral as to whether the notion of sameness of meaning on which the meaning attributions are based is purely factual. See also Section 6 of the Postscript to that chapter.)

6. So far I have simply defended the *possibility* of a linguistic theory of meaning attributions, and said that it offers a neat answer to the question of why the instances of the meaning schemas don't seem open to doubt on empirical grounds. I don't claim that the linguistic theory is inevitable.

One rather natural alternative to the linguistic theory is that  $\rangle$ that= clauses and other bracket-terms are not denoting terms:  $\rangle$ believes that= and  $\rangle$ means that= and  $\rangle$ means  $\langle \rangle$ = are to be viewed as operators that convert sentences and other expressions into predicates, with the component expression more or less semantically inert. Stephen Schiffer has pointed out to me that there are inferences in which  $\rangle$ that= clauses *seem* to figure as singular terms: e.g. the inference from  $\rangle$ Sue believes that  $E=mc^2=$  and  $\rangle$ Einstein's theory is that  $E=mc^2=$  to  $\rangle$ Sue believes Einstein's theory=. But one could also validate the inference by supposing that the first premise was short for  $\rangle$ Sue believes the theory that  $E=mc^2=$  and the second premise was short for  $\rangle$ Einstein's theory is the theory that  $E=mc^2=$ . This does save the view of  $\rangle$ that= clauses as non-denoting, but of course it requires that belief sentences can take singular terms as well as  $\rangle$ that= clauses as complements, and so we still must decide whether those singular terms denote linguistic expressions or intentional entities. The non-denoting operator view just defers the question.

The other alternative we need to consider is that  $\rangle$ that= clauses, and bracket terms more generally, refer to some sort of intentional entities: propositions, concepts, and the like, taking concepts

not to be components of mental states, but to be intentional entities assigned to such components. (I will occasionally use the term >meaning= as a catchall for properties, concepts and the like; as remarked earlier, this is somewhat idiosyncratic usage in the case of sentences with indexicals, for it involves taking two tokens of >She is beautiful= said of obviously different people as being assigned different meanings@. Frege=s term >sense= or Kaplan=s term >content= might be better, but I trust that no confusion will ensue.) I am not strongly opposed to this intensionalist proposal, as long as the intentional entities are construed in a sufficiently bland way. For instance, it is certainly possible to introduce, for each expression *e* of our language (or rather, each use of a such an expression) a corresponding intentional entity <sup>i</sup>*e* (the intentional entity corresponding to *e*) with no particular properties other than those induced from the properties of *e*; more or less in the way that Frege introduced directions on the basis of lines.<sup>2</sup> (As Stephen Schiffer once nicely put it: propositions should be viewed simply as shadows of sentences, and concepts simply as shadows of certain subsentential expressions.) The theory that introduces these entities can be set up so that it conservatively extends the underlying theory, in which case it is hard to see how it can be regarded as seriously problematic.<sup>3</sup> The danger of doing so, however, is that we will start talking about propositions, concepts, etc. in ways not licensed by this method of introducing them. Notice for instance that this method of introducing propositions would seem compatible with viewing them as not importantly different from either (i) classes of synonymous sentence-tokens (on the conception of synonymy introduced before, on which tokens of >She is beautiful= said of obviously different people are non-synonymous); or (ii) equivalence classes of meaning characteristics (as defined in Section 2). Either of these would make the propositional view basically just a notational variant of the linguistic view (though the second involves a commitment to the *not* radically Quinean@ linguistic view (LV), as opposed to (QLV)). So to assume that propositions are not to be viewed as simply classes of synonymous sentences or as classes of meaning characteristics is to assume something that goes beyond what is licensed by the Fregean introduction technique, and the intensionalist would seem to need an argument for the further assumption. But having made this point, I will not press it: from here on I will concede to the intensionalist that propositions are not just classes of linguistic entities.

My main qualm about talking in terms of propositions and the like isn't that doing so is *incorrect*, but that it tends to obscure the fact that the explanatory role of propositions derives from that of the sentences that they shadow@. I discussed this in Section 4.

Moreover (as Davidson 1967 observed) it seems impossible to develop any nontrivial theory of these intentional entities. For instance, how do they combine? The view of meaning-combination that

seems most in line with the ordinary notion of meaning is that sameness of meaning requires being built up in the same way from synonymous parts. In that case, the approach requires the introduction of a concatenation operation on meanings, analogous to that on expressions. Given that, we could of course give a simple inductive clause for correlating expressions with meanings: for any expressions  $E_1$  and  $E_2$ , if  $E_1$  means  $m_1$  and  $E_2$  means  $m_2$  then  $E_1 \wedge E_2$  means  $m_1 \wedge m_2$ . How exciting!<sup>4</sup>

We could extend this *Atheory@*, by connecting sameness of meanings with sameness of meaning characteristics (in the non-intentional sense introduced in Section 2): two expressions have the same meaning iff they have equivalent meaning-characteristics. But even with such an extension, the theory as just developed makes no connection between these meanings on the one hand and truth-conditions and reference-conditions on the other: that needs to be added. Nothing tells us, for instance, that the meaning associated with  $\text{>Plato=}$  picks out Plato. We must add some sort of theory of reference (perhaps a disquotational theory) to get this. (It won't help to somehow build into the notion of  $\text{'Plato'}$  that no intentional entity can be  $\text{'Plato'}$  unless it picks out Plato. For if we did so, it would no longer be trivial that the meaning associated with  $\text{>Plato=}$  is  $\text{'Plato'}$ : we would need an explanation of why the meaning of  $\text{>Plato=}$  is  $\text{'Plato'}$ , and it isn't obvious why the claim that the meaning of  $\text{>Plato=}$  is  $\text{'Plato'}$  should be beyond reasonable doubt.)

And we must add similar *Atheories of reference@* for other expressions, e.g. predicates and logical symbols; for instance, a theory that tells us that any expression whose meaning is  $\text{'or'}$  picks out the appropriate truth-function. We also must add a theory of truth, showing how the meanings of complex expressions contribute to truth conditions; though presumably this will reduce to the usual inductive account of truth in terms of reference, plus the account of reference. And isn't all the interesting work now being done by the theories of reference and truth, plus the account of which equivalence relation on the set of meaning characteristics underlies translation? The intentional meanings seem to be completely idle. Note for instance that the inductive specification of the meanings has no real base clause other than the base clause for reference. You can't specify a nontrivial base clause by conjoining all sentences of form  $\text{>e= means 'e'@}$  for atomic  $e$ , because we have given no content to the phrase  $\text{'e'}$  other than *Athe actual meaning of >e= in our own language@*, the contemplated base clause adds nothing. The only substantial base clause available is the conjunction of sentences like *Aan expression that means 'Plato' refers to Plato@*, *Aan expression that means 'or' obeys the truth-table DISJUNCTION@*, and so forth. So the compositional theory can't say anything about how the meanings associated with coreferential atomic expressions differ (unless at least one of them is

synonymous with another expression), it can only say that they differ somehow.

The danger that the introduction of intentional entities will appear more explanatory than it really is becomes especially acute if we write our theories of truth and reference in the form  $(T_G)$  and  $(R_G)$  used at the end of the previous section. I remarked there that on a linguistic view of meaning attribution, these are just extensions of  $(T)$  and  $(R)$  to other languages via sameness of meaning; the central role of our own language is still easily seen, because of its role in the meaning attributions used in  $(T_G)$  and  $(R_G)$ . But if one introduces propositions and other intentional entities, and takes  $\lambda$ that= clauses and other bracket terms to refer to them, then this central role of our own language is easily blurred. Indeed there can then be a real unclarity as to whether  $(T_G)$  and  $(R_G)$  express a **A**deflationary@view of truth not essentially different from disquotationalism, or a thoroughly **A**inflationary@view of truth, or something in between. The view is deflationary as long as **A**the proposition that Plato taught Aristotle@is interpreted as simply meaning

(A) the proposition actually meant by our sentence  $\lambda$ Plato taught Aristotle=, for then the theory of truth  $(T_G)$  builds in the fact that  $\lambda$ Plato taught Aristotle= as we actually use it is true if and only if Plato taught Aristotle. The presence of that in the theory of truth (as opposed to, in a result that we hope to arrive at from the theory of truth together with the facts about the use of our language) is disquotationalism. If on the other hand (i) **A**the proposition that Plato taught Aristotle@is interpreted as meaning something like

(B) the set of possible worlds in which Plato taught Aristotle  
or

(B=) the ordered triple  $\langle$ Plato, Aristotle, the relation of teaching $\rangle$ ,

and (ii) we think of semantic relations as connecting our own expressions to propositions and their components in just the way that they connect foreign expressions to propositions and their components (our own language being viewed as in no way special), then the view is thoroughly inflationist. (Note that even on a view of propositions as sets of possible worlds, one could take **A**the proposition that Plato taught Aristotle@as meaning either (A) or (B), and that this makes a crucial difference to the interpretation of the intensionalist version of  $(T_G)$ ; analogously for a view of propositions as Russellian.)

I don't mean to be arguing against a thoroughly inflationist view of truth and reference, I am simply saying that we shouldn't smuggle one in by unclear use of terms like  $\lambda$ proposition= and  $\lambda$ concept=, and that real care is required to avoid doing so.

But if one keeps in mind the explanatory idleness of meanings (propositions, concepts, etc.),

and is clear enough in the use of phrases like >the proposition that Plato taught Aristotle= to avoid inviting confusion, then I think introducing such talk is harmless. And if we indulge in such talk, we can understand attributions of meaning, belief etc. in terms of propositions and the like.

One might suppose that Quinean considerations count against this. Quine noted in some of his early writings that the notion of synonymy is much clearer when restricted to words or sentences in the same idiolect than it is in the inter-idiolect case. If someone accepts >The Son of Sam was a serial killer= but not >David Berkowitz was a serial killer=, or even if he merely can imagine evidence that would lead him to accept one but not the other, that's enough to show that >The Son of Sam= and >David Berkowitz= don't mean the same thing for that person. This suggests that intra-idiolect sameness of meaning ought to be explainable by some sort of substitutivity criterion: without trying to be very precise, the idea is that two terms of the same idiolect mean the same if substitution of one for another inside a sentence (except in certain special contexts like quotation marks or sentential attitude constructions) doesn't affect the speaker's epistemic attitudes toward the sentence; for instance, it doesn't affect what observations or what other sentences count as supporting the sentence. Terms that are synonymous within one idiolect of a language tend to be synonymous within other idiolects of the language as well, so it does little harm to speak of intra-*linguistic* synonymy. But substitutivity criteria obviously can't work in the case of inter-linguistic synonymy.<sup>5</sup> (Such criteria don't really work even for inter-idiolect synonymy, though there we can stipulate that smooth communication under the identity translation should suffice for synonymy.) And Quine thought that nothing else would work in the inter-linguistic case either: the distinction between differences of meaning and differences of belief can be given no principled basis. We can still invoke translations from one language to another, but shouldn't view them as attempts to reflect a pre-existing synonymy relation.

It isn't entirely clear that the considerations Quine adduced show more than the indeterminacy of inter-linguistic synonymy. If that is all it shows, then of course there is no threat to talk of propositions and the like, provided that we recognize an indeterminacy in that notion. But Quinean considerations are often thought to cut deeper than this: to show that @there is no synonymy relation that translation is an attempt to reflect@. (I don't say that this wears its meaning on its face, but I think it probably can be made sense of. One possible approach to doing so was suggested in Chapter 4, note 27. And one reason why someone might want to maintain such a view will be mentioned in Chapter 10, note 2.) But even this is compatible with the innocent introduction of intentional entities, as long as one construes those entities as *local entities*. That is, we take each language to have its own supply of meanings,

not objectively comparable to the meanings of any other language. (If you like, you could think of the meanings of the expressions of language L as just equivalence classes of meaningful L-sentences under the equivalence relation of intra-linguistic synonymy; though this construal is no more essential to the idea than the analogous construal when we were using a more general synonymy relation.) We could take it to be an objective matter which of the  $L_1$ -meanings an  $L_1$ -sentence means, and which of the  $L_2$ -meanings an  $L_2$ -sentence means, but  $L_1$ -meanings and  $L_2$ -meanings are simply distinct entities; a mapping of  $L_1$ -meanings to  $L_2$ -meanings can be suitable or unsuitable relative to certain goals, but talk of objective correctness makes no sense here. This would allow us whatever convenience meanings as entities might be thought to have (e.g. in the semantics of belief sentences), without any commitment to views about interpersonal comparison that might be deemed suspect.

Indeed, the Fregean analogy with directions has an illuminating extension to this case. Let's pretend that space as a whole is highly non-Euclidean, but that it has a number of large Euclidean regions within it. Within each Euclidean region, geometry is as you're used to. In particular, within each region the notion of parallelism of line segments makes clear sense, and is an equivalence relation. So within each region we can assign to a line segment a direction: a direction is simply what all parallel line segments in the region have in common. We can do this within each of the flat regions. Now, a naive philosopher might say,

AOK, within region 1 we can assign directions to line segments, and sensibly ask whether two line segments have the same direction; and within region 2 we can also assign directions to line segments, and sensibly ask whether two line segments have the same direction. So it ought to make sense to ask whether a line segment in region 1 has the same direction as a line segment in region 2.@

But in fact it doesn't make sense: if you take a line segment in region 1 and transport it to region 2, keeping it as straight as possible along the journey, how it ends up aligned in region 2 will depend on the path of transport. (You can see this in two dimensions by taking the regions to be small flat regions on the surface of the earth, one on the equator and one at one of the poles.) Directions are perfectly well behaved local entities, in that within each flat region we can assign local directions in a completely consistent manner; but the local directions for one region aren't straightforwardly comparable to those for another, they are comparable only relative to a method of transport. On a strong understanding of Quine's indeterminacy thesis, meanings can be viewed as just like that: they make pretty good sense locally (i.e. within a language, or maybe just within a speaker), but they aren't straightforwardly

comparable inter-linguistically, but rather, are comparable only relative to a method of translation.<sup>6</sup>

Whether this is a good picture is not something I want to argue here. My only claims are (i) that as long as one is careful enough in talking of propositions, we can accommodate the Quinean picture; and its flip side (ii) that we need to be careful not to smuggle in the assumption that the Quinean picture is wrong by uncritically assuming that the propositions one has innocently introduced in analogy with directions are globally defined. This fits with the general picture that propositions and other intentional entities are things that we can take or leave as we see fit: we don't need them for attributions of meaning or belief or the like, but as long as this is clearly recognized, and we avoid certain confusions of ambiguity, they do no harm.

7. Early in the paper I noted that the meaning schema (M) seems to have a special epistemological status: its instances are not open to doubt on empirical grounds. And I said that the linguistic view of meaning attributions gives one possible explanation of this. Are there other explanations? In this final section I'll consider several such.

It might be thought that there is a completely obvious explanation of the special status of (M). It involves the idea that to understand an expression is to know its meaning. I should note in passing that this is ambiguous between two interpretations. One is that to understand an expression is to know what it means for me; the other is that to understand an expression is to know its meaning in the public language. These two interpretations should not be viewed as giving conflicting theses about understanding, they should be viewed as showing an ambiguity in the notion of understanding. In one sense I can understand a term even if my understanding doesn't coincide with other people's; in another I can't. Given the focus in this paper on the idiolect of an agent, it is the former sense that is the more relevant. So: for me to understand an expression is for me to know what it means for me.

But then isn't the special status that the schema (M) has for me quite trivial? After all, the instances of the schema that I can be said to *believe* are just those where the term  $\langle \text{grug} \rangle$  in question is one that I understand. If I have no understanding of  $\langle \text{grug} \rangle$ , then I won't accept

$\langle \text{grug} \rangle$  means  $\langle \text{grug} \rangle$ .

(And even if I did it wouldn't count as a belief.) But for me to understand  $\langle \text{grug} \rangle$  is just for me to know what it means, so of course I am going to know the meaning of all the terms I understand! Where's the mystery, or the need of the linguistic theory of meaning attributions?

I think this is mistaken. It's worth noting that while the claim that to understand an expression is

to know its meaning is in a way unexceptionable, it may also be misleading. What is unexceptionable is that for an ordinary person (one who has the concept of meaning), understanding a term goes along with knowing what it means; one understands  $\succ$ metacompact $\Leftarrow$  if and only if one knows that it means  $\langle$ metacompact $\rangle$ . What would be misleading is to take for granted that this knowledge that  $\succ$ metacompact $\Leftarrow$  means  $\langle$ metacompact $\rangle$  *explains* one's understanding of  $\succ$ metacompact $\Leftarrow$ . In my opinion, that would be getting things backwards. Understanding a term (in the idiolect sense of  $\succ$ understanding $\Leftarrow$ ) involves, roughly, having a sufficiently rich conceptual role for it.<sup>7</sup> It isn't a matter of knowing that p, for any p whatsoever, let alone a p about the meaning of the word. It is the understanding of  $\succ$ metacompact $\Leftarrow$ , together with one's acceptance of all instances of the schema (M), which explains the fact that one accepts

$\succ$ metacompact $\Leftarrow$  means  $\langle$ metacompact $\rangle$ ,

(and the fact that accepting it suffices for believing it). And in conjunction with the impropriety of revising (M) on empirical grounds, this yields the conclusion that the instances of (M) are known and empirically unrevisable. But of course this presupposes the special epistemological status of (M): it presupposes that the instances of (M) should be accepted whatever the empirical evidence. And it is that special status that was supposed to be explained.

So at best, the proposed explanation of the special status of (M) is one that works only on a very contentious (not to say ridiculous) version of the connection between meaning and understanding: a version on which propositional knowledge about meaning plays an explanatory role in understanding. But even on such a view of the connection between meaning and understanding, the explanation of the special status of the instances of (M) is faulty. For all it does is  $\Leftarrow$ explain $\Leftarrow$  the generalization

For any expression, if we understand it then we know what it means  
(by reducing this generalization to a tautology). That together with the fact that  $\succ$ metacompact $\Leftarrow$  means  $\langle$ metacompact $\rangle$  gives us a trivial explanation of

If we understand  $\succ$ metacompact $\Leftarrow$  then we know that it means  $\langle$ metacompact $\rangle$ .

But what was asked for wasn't this, but an explanation of the consequent. Well, we'd have that if we had an explanation of the antecedent. But on the view of understanding in question, we could only explain the antecedent by explaining our knowledge of what  $\succ$ metacompact $\Leftarrow$  means: viz., that it means  $\langle$ metacompact $\rangle$ ! So even on this contentious view of understanding, the  $\Leftarrow$ explanation $\Leftarrow$  of the status of (M) is completely circular.

I turn now to a second attempt to explain the special status of (M). The idea of the approach is

to note that the instances of (M) are built into the notion of meaning, in the following sense: it is part of learning to use the word ›means= that we come to accept all instances of (M), and our acquiring this body of beliefs is central to the notion of meaning serving the purposes it serves. Its being part of the meaning in this sense is all we need to legitimate the practice of giving default status to all instances of (M); only an absurdly foundationalist epistemology would require more. And nothing in this explanation commits one to the linguistic view of meaning attributions.

As an explanation of the default status of the instances of (M), I think this approach is unexceptionable. But what I asked for was more: an explanation of why we should regard the instances of (M) as empirically indefeasible, that is, of why we should think there can be no empirical evidence against the instances of (M). To regard (M) as empirically indefeasible is much stronger than merely to give it default status.

There are many examples to illustrate this: consider for instance ›temperature=, or (better for my purposes) the comparative predicate ›has higher temperature than=. I take it that a few hundred years ago, part of the ordinary mastery of that predicate was the acceptance of such beliefs as that if one body felt substantially warmer than another to a normal observer in normal conditions then its temperature was higher. Such beliefs would have been taken to be central to the explanatory purposes of the notion of having a higher temperature. These beliefs were in a perfectly good sense built into the meaning of ›has a higher temperature=, and had a default status in that no one was required to argue for them on empirical grounds. But these beliefs were not empirically unrevisable: once people started exploring the physical underpinnings of temperature and of feelings of warmth, it was discovered that the beliefs were false, in that things other than temperature-differences (for instance, differences in thermal conductivity) play a substantial role in determining feelings of comparative warmth. In the case of ›means= however this cannot happen: if one insists on construing the claims that ›dog= means ‹dog› and that ›metacompact= means ‹metacompact› as claims about ›dog= bearing the meaning relation to <sup>i</sup>dog<sup>i</sup> and ›metacompact= bearing the meaning relation to <sup>i</sup>metacompact<sup>i</sup>, then these claims about the extension of the meaning relation seem empirically unrevisable. So our having learned to accept the instances of (M) as central to the meaning of ›means= is not by itself enough to explain their immunity to empirical revision.

I think it is fairly clear that explaining the empirical unrevisability of (M) requires a somewhat Adeflationary@ attitude toward meaning attributions to our own terms. The reason that our views about the relation of *having higher temperature than* can be empirically revised even when Abuilt into@

the meaning of ›temperature= has to do with the fact that ›temperature= is an explanatory concept. If the attribution of intentional entities (meanings) to our own terms isn't taken to be explanatory, there is no mystery as to how it can be reasonable to hold onto such attributions whatever empirical discoveries we might make.

Of course there is no doubt that attributing what I've called *meaning characteristics* (conceptual roles, indication relations, etc.) to expressions is explanatory: that is explanatory in a straightforward causal sense. And attributions of intentional entities to expressions can be used in explanations, in so far as the attribution of an intentional entity is used simply as a way of (perhaps vaguely) picking out the relevant meaning characteristics. But the way in which the ascription of an intentional meaning picks out meaning characteristics makes use of the fact that intentional meanings are mere shadows of expressions. That is, if we explain a person's uttering (or silently believing) a certain German sentence on the basis of its meaning 'there have been many British kings', the intentional entity we've picked out in the explanation has been picked out via our expression ›there have been many British kings=; our explanation amounts to the claim that the person's utterance (or inner state) has meaning characteristics sufficiently like those actually possessed by our sentence ›there have been many British kings= for the purposes at hand. Nothing about the relation between expressions and meanings, other than the fact that sameness of meaning goes by sameness in type of meaning characteristic, plays any role in this explanation; or in any other legitimate explanation, so far as I can see. It is no wonder that if meanings are viewed as simply shadows of our expressions, and the shadowing relation is not put to explanatory work, then there can be no empirical reason to revise the meaning schema. But if intentional meanings are only shadows of expressions, then they would seem to be entities that we can take or leave as we see fit. And if intentional meanings have a more thoroughly explanatory role than this, then it is hard for me to see why the special epistemological status of (M) isn't a mystery.<sup>8</sup>

## Footnotes

---

1. Stephen Leeds (1979) gives a superficially quite different response to the Church argument, which may not ultimately be as different as it initially seems. A key part of his response is that the German word ›bedeuten= doesn't literally mean the same as ›means=, or even have the same extension: the first relates ›dog= and ›Hund= to ›Hund= but not ›dog=, the second relates them to ›dog= but not ›Hund=. But

---

›bedeutet= and ›means= do nonetheless stand in an important homology relation, similar to that between the extensionally different predicates ›the temperature-in-Fahrenheit of x is r= and ›the temperature-in-Centigrade of x is r=, which makes it reasonable to translate ›bedeutet ›Hund= as ›means ›dog=. Leeds, like me, insists that capturing the literal content is simply not important in translation of ›means= sentences (in his case, it's because we have no way to translate that preserves literal content). Indeed, I see no reason to think that Leeds and I differ as to the ›quasi-content@ that should be preserved. The accounts of quasi-content in terms of literal content differ, but that needn't reflect a difference about quasi-content, but merely the difference in the views about literal content: I'm assuming the Church-Schiffer view, according to which ›means= and ›bedeutet= literally have the same content, while he is assuming an alternative.

Other comparisons: The fact that translators merely ›quasi-translate@ was also emphasized by Dummett (1973, p. 372); he too says that this undermines the Church argument. It is mostly because of the added twist in Schiffer's argument that I saw fit to give more detail. Sellars (1962) used his dot quotes (mentioned above) for attribution of meanings and attitudes; if I'm right in saying that his dot quotes were to name intersubjectively ascribable inferential roles, this isn't a linguistic view (nor is it an adequate view, in light of standard externalist objections to sameness of inferential role as a criterion of translation).

2. I use these new <sup>i</sup>-quotes instead of the earlier <-quotes because my use of the latter is neutral to what if anything the <-terms refer to. <sup>i</sup>-quotes and #-quotes are competing accounts of <-quotes.

3. I should add that a large part of the reason that conservative extensions of a theory are unproblematic is that it is possible to construe the added entities as fictitious, or convention-dependent; we can talk in terms of such entities without really being platonist about them. But I will not press this point.

4. We could also include definitions of the types of meanings, which would entail e.g. that if  $m_1$  is a 1-place predicate concept and  $m_2$  an individual concept then  $m_1 \wedge m_2$  is a proposition. This doesn't add much to the excitement that the theory induces.

5. This is certainly obvious when there are no bilingual speakers. And the presence of bilinguals doesn't change anything: the first bilingual's choice of a mapping from one language to the other is partly arbitrary; if other bilinguals were to make a choice independent of what the first bilingual has done, there is no reason to think they would make the same choice.

6. Note that in the geometric analogy, it is natural to say not merely that it is indeterminate whether line segments in different regions are parallel, but that the question of their parallelism makes no sense except relative to a method of transport. I am inclined to say the same in the synonymy case, though in both cases the matter deserves more discussion.

7. Understanding it in the more social sense involves, roughly, having a sufficiently rich conceptual role

---

for it that accords in the appropriate way with the conceptual role that others have given it.

8. Stephen Schiffer's skeptical reaction to the previous chapter provided the main impetus for writing this, and he made many useful comments along the way; Paul Horwich and Brian Loar gave me comments on a previous draft that led to a major change in emphasis. Ken Akiba, David Barnett, Jared Blank, Ray Buchanan, David Enoch, Joshua Schechter, Brad Skow and James Woodbridge all made comments that significantly influenced the final version of this chapter and/or the related parts of the postscript to the previous chapter.