CHAPTER 23.

RELEVANCE, TRUTH AND MEANING

Graham Priest and Jan Crosthwaite

1. Introduction. A central topic in current philosophy of language is the role of truth in the theory of meaning. In particular, many people hold that at the core of a theory of meaning for a language is a Tarski-type truth theory,[1] i.e. an axiomatic theory which entails, for every sentence s of the language in question, a theorem of the form

s is true if and only if p

where p is the translation of s in the language being used.[2] The 'if and only if' in this theorem, the T-scheme, is usually taken to be the material bi-implication '≡'. This is not part of the idea as such. As Davidson puts it:

> Convention T, in the skeletal form I have given it, makes no mention of extensionality, truth functionality, or first order logic. It invites us to use whatever devices we can contrive appropriately to bridge the gap between sentences mentioned and sentences used (73, pp.78-9).

This may be a surprise to those who think of Davidson as working *essentially* in the Quinean research programme, aiming to give an account of meaning which avoids the Quinean *bête noire* intensionality. However the thrust of the Quinean arguments used in Davidson 67 is against meanings *as entities*. It is this which separates Davidson from Frege, Montague, *et al.*[3]

Thus, the use of classical extensional logic gets into the Davidsonian programme on the coat-tails of a general acceptance of "classical" logic. However, there are many and cogent arguments against formalising 'if and only if' in this way.[4] A much better formalisation of the English connective is a genuine relevant bi-implication '$\longleftrightarrow$'. In this paper, we will argue that a number of the difficulties of this account of meaning accrue to it solely in virtue of its use of an irrelevant logic and that therefore the use of a relevant logic is to be highly recommended.

We will take up these issues in sections 3-6; but before we do so there is an important technical issue to be settled. In virtue of the *de facto* assumption that classical logic is an integral part of this kind of approach to meaning, it might be thought that a theory of truth must use classical logic, that such a theory cannot be built on relevant logic. This is false. The properties of implication that are actually used in proving instances of the T-scheme are minimal. In particular, it is possible to give a truth theory for a first order language (which itself contains a relevant $\rightarrow$) in a "relevant metalanguage". Anyone to whom this is clear, or who is content to take our word for it can move straight on to section 3.

J. Norman and R. Sylvan (eds.), Directions in Relevant Logic, 377–397.

2. Relevance and the T-scheme. Let us take a first order language O, though for good measure we will add a genuine implication operator '⇒'. The vocabulary of O is as follows:

variables: $v_0, v_1, v_2, \ldots$
n-place predicates: R^n_i, $i \in I_n$ (where for each n, I_n is a set of indices).
connectives: &, ~, ⇒.
quantifier: Σ.
brackets: [,].

Formation rules are as usual, and disjunction and universal quantification can be thought of as defined in the usual way.

The truth theory of O will be given in a metalanguage M. It will pay us to be reasonably precise about M. M is a first-order language with function symbols and constants. Its vocabulary is:

variables: $a_1, a_2, a_3, \ldots$
constants: $\underline{\alpha}$ for every variable and predicate symbol, α, of O.[5]
function symbols: Ap, Sub, Neg, Con, Imp, Ext, Pred^n (all n).
predicates: Sat, Q^n_i (all $i \in I_n$, all n).
connectives: $\wedge, \vee, \neg, \rightarrow$.
identity: =.
quantifiers: $\forall, \exists$.
brackets: (,).

Intuitively, the function symbols are to be understood as follows: The interpretation of Pred^n is an n+1-place function which, applied to an n-place predicate of O, α, and the variables of O, $\beta_1 \ldots \beta_n$, gives the formula of O which is α followed by $\beta_1 \ldots \beta_n$. If we think of this as the n+1 tuple $<\alpha, \beta_1 \ldots \beta_n>$ then we may take this to be the value of the interpretation of Pred^n whatever $\alpha, \beta_1 \ldots \beta_n$ are. The interpretation of Con is a two-place function which applied to formulas of O, α, β, gives their conjunction, which we may take to be the 5-tuple $<(, \alpha, \&, \beta,)>$ thus defining the interpretation of Con universally. Similar remarks apply to Neg, Imp, and Ext, for negation, implication and existential quantification respectively.

The interpretation of Ap is the two-place function of functional application. However, in case its first argument is not a function, or its second argument is not in its domain, we define it slightly more generally as the function f such that

$$f(xy) = \begin{cases} z \text{ if } z \text{ is the unique } w \text{ such that } <y,w> \in x \\ y \text{ otherwise} \end{cases}$$

The interpretation of Sub is the three-place function g such that g(xyz) is the function which is the same as x except that its value at argument y is z. However since, again, x may not be a function, or y may not be in its domain, we define it more generally thus:

$$g(xyz) = (x - \{w | \exists u \; w = <y,u>\}) \cup \{<y,z>\}$$

The predicate Sat is a two-place predicate such that Sat $\alpha\beta$ is thought of as 'α satisfies β'. Q_i^n is an n-place predicate. The formation rules of M are the obvious ones consistent with the above remarks.

To ease notation in the metalanguage we will accept the following conventions:

$\alpha \longleftrightarrow \beta$ is $(\alpha \rightarrow \beta) \wedge (\beta \rightarrow \alpha)$
$\alpha(\beta)$ or α_β is Ap$\alpha\beta$
$\alpha(\beta/\gamma)$ is Sub$\alpha\beta\gamma$

Where it does not matter which M-variable is to be used we will write a, b, ... s, s', ... x, y, ... (possibly with subscripts), reserving symbols in the second group for places where we are primarily thinking of the variables as denoting a function whose domain is the variables of O, and symbols in the final group for places where we are primarily thinking of the variables as denoting a symbol or string of symbols of O. (However, this is only for ease of reading. The official theory is single-sorted.)[6]

The underlying logic of the metalanguage is some predicate extension of a relevant logic, with identity. Virtually any relevant logic will do, for example that of Priest 80. The only properties of bi-implication which are really necessary are transitivity and the principle

$$\alpha \longleftrightarrow \beta \;/\; \gamma \longleftrightarrow \gamma^\alpha_\beta \qquad (\Pi)$$

where γ^α_β is γ with occurrences of α replaced by β (subject to the usual restrictions on free variables). The only axioms we need other than those of the truth theory proper are:

i) $\underline{v}_i \neq \underline{v}_j$ if i and j are different
ii) Ap s(x/a) x = a
iii) x $\neq$ y $\rightarrow$ Ap s y = Ap s(x/a) y

which are all true under the interpretation of Ap and Sub given.[7] In the more general setting of a (relevant) set theory these axioms would be provable in the usual way.

The machinery of an appropriate metatheory can be set up in many different ways. We have chosen the above way to make the theory of truth for O and the proof of the T-scheme as simple as possible. Indeed, as we shall see, the truth theory itself is quite straightforward and orthodox. The axioms of the theory are the obvious:

1) Sat s $\text{Pred}^n \underline{R}^n_i x_1 \ldots x_n \longleftrightarrow Q^n_i s(x_1) \ldots s(x_n)$, all i ε I_n, all n
2) Sat s Con x y $\longleftrightarrow$ (Sat s x $\wedge$ Sat s y)
3) Sat s Neg x $\longleftrightarrow$ $\neg$Sat s x
4) Sat s Imp x y $\longleftrightarrow$ (Sat s x $\rightarrow$ Sat s y)
5) Sat s Ext x y $\longleftrightarrow$ $\exists$a Sat s(x/a) y

To prove the truth scheme, three steps are necessary. First we must specify for each formula of O, ϕ, a term of M, $\underline{\phi}$, which is its (canonical) name. This is done recursively in the usual way. The basis clauses are of the form:

$\underline{R}_i^n \underline{v}_1 ... \underline{v}_n$ is Pred^n $\underline{R}_i^n \underline{v}_1 ... \underline{v}_n$,

and recursive clauses such as:

$\underline{\exists v_i}$ $\underline{\phi}$ is Ext $\underline{v}_i$ $\underline{\phi}$

then generate names for all formulas.

Secondly, we must define the translation of an O-sentence into M, with respect to a sequence s. If ϕ is any sentence of O, its translation into M with respect to s, $(\phi)_s$, is the M-sentence formed by replacing every free variable 'v_i' by '$s(\underline{v}_i)$', every predicate 'R_i^n' with 'Q_i^n' every bound variable 'v_i' (including those occurring immediately after 'Σ') with 'a_i', every '&' with '$\wedge$', every '$\sim$' with '$\neg$', every '$\Rightarrow$' with '$\rightarrow$', every 'Σ' with '$\exists$', every '[' with '(' and every ']' with ')'.

The third, and main, step is to prove the satisfaction scheme in M:

$$\text{Sat s } \underline{\phi} \longleftrightarrow (\phi)_s$$

where ϕ is any formula of O. This is proved by the usual recursion over formation.

The basis is where ϕ is $R_i^n v_{k_1} \ldots v_{k_n}$. Here we need to prove that

$$\text{Sat s Pred}^n \underline{R}_i^n \underline{v}_{k_1} \ldots \underline{v}_{k_n} \longleftrightarrow Q_i^n s(\underline{v}_{k_1}) \ldots s(\underline{v}_{k_n})$$

and this is a simple instantiation of 1). The cases for &, $\sim$ and $\Rightarrow$ are all similar. Here is the case for $\sim$.

Suppose that	Sat s $\underline{\phi}$	$\longleftrightarrow (\phi)_s$.
Then by (Π)	$\neg$Sat s $\underline{\phi}$	$\longleftrightarrow \neg(\phi)_s$
i.e.	$\neg$Sat s $\underline{\phi}$	$\longleftrightarrow (\sim\phi)_s$.
But by 3) and the transitivity of bi-implication		
	Sat s Neg $\underline{\phi}$	$\longleftrightarrow (\sim\phi)_s$
i.e.	Sat s $\underline{\sim\phi}$	$\longleftrightarrow (\sim\phi)_s$ as required.

The final case is for $\exists$. This is proved as follows.

$$\text{Sat x Ext } \underline{v}_i\, \underline{\phi} \longleftrightarrow \exists a_i \text{ Sat } s(\underline{v}_i/a_i)\, \underline{\phi} \quad \text{from 5)}$$
$$\longleftrightarrow \exists a_i\, (\phi)_{s(\underline{v}_i/a_i)} \text{ by Induction Hyp.}$$

Now if we can prove $\exists a_i(\phi)_{s(\underline{v}_i/a_i)} \longleftrightarrow (\exists v_i \phi)_s$, ($\Omega$), we are home. Let us write t for $s(\underline{v}_i/a_i)$. Then $(\phi)_t$ and $(\phi)_s$ are identical except that where ϕ contains v_j free, $(\phi)_t$ contains $t(\underline{v}_j)$ and $(\phi)_s$ contains $s(\underline{v}_j)$. Now consider $(\phi)_t$ and let v_j be any variable that occurs free in ϕ except that j is different from i. Then, by i) and iii)

$$t(\underline{v}_j) = s(\underline{v}_j).$$

Now by applying the scheme

$$a = b \;/\; \alpha \longleftrightarrow \alpha^{a}_{b}$$

the substitutivity of identicals, to all such terms, with α as $(\phi)_t$, we obtain a formula ψ which is provably equivalent to $(\phi)_t$, and which is the same as $(\phi)_s$ except that where $(\phi)_s$ contains $s(\underline{v}_i)$, ψ contains $t(\underline{v}_i)$. But $t(\underline{v}_i) = s(\underline{v}_i/a_i)\ (\underline{v}_i) = a_i$ by ii). Substituting this identity in ψ we obtain a formula ψ' which is provably equivalent to ψ and is the same as $(\phi)_s$ except that where $(\phi)_s$ contains $s(\underline{v}_i)$, ψ' contains a_i. By (Π)

$$\exists a_i(\phi)_t \longleftrightarrow \exists a_i\psi'.$$

But $\exists a_i\psi'$ is exactly $(\exists v_i\phi)_s$. Hence we have proved (Ω).

The final step in proving the truth schema is now simple. 'x is true' is defined in the usual way as '∀s Sat s x'. Now let ϕ be any closed formula of O. We know that

$$\text{Sat s } \underline{\phi} \longleftrightarrow (\phi)_s.$$

But since ϕ is closed 's' has no free occurrences in $(\phi)_s$. Thus

$$\forall s\ (\phi)_s \longleftrightarrow (\phi)_s$$

and it follows that

$$\underline{\phi} \text{ is true} \longleftrightarrow (\phi)_s$$

(where $(\phi)_s$ does not, of course, depend on s).

3. The Problems of a Material T-scheme in a Theory of Meaning. We have established that there is no problem in constructing a truth theory in which the bi-conditional of the T-scheme is relevant. We will now review the aspects of a Davidsonian theory of meaning which are germane to our purposes, and state three problems that classical logic, and in particular a material bi-conditional in the T-scheme, produces. In subsequent sections we will discuss each problem in more detail and show how a relevant bi-conditional solves the problem.

Let us call the instances of the T-scheme, "T-sentences". Then our central concern is how T-sentences are supposed to function in a theory of meaning for a language. The Davidsonian view is that it is precisely the T-sentences of a truth theory which play the crucial role of providing knowledge sufficient for interpretation of that language.[8] (One need not know all the truth theory in order to use it to interpret, only the T-sentences.[9])

We can now state the three problems we have in mind. First, there is a problem with the notion of T-sentence. The original specification of T-sentences (which we gave in section 1) uses the notion of translation. However, explicit use of this notion is not available to someone who wishes to take a theory of truth to provide an account of meaning and meaning-related notions for a certain language. To invoke a notion of translation would presuppose that a) the meanings of object-language sentences were known and b) the notion of translation (= meaning preservation) was clear. Of course, Tarski's original interest was in characterising *truth*. Hence he could legitimately presuppose these notions. However, this is no longer the case if we wish to use the theory of truth as a theory of meaning.

The second problem follows from the fact that this proposal for giving meanings is supposed to work for natural languages. For this raises the hoary old problem of the semantic paradoxes. Tarski, of course, saw the semantic paradoxes as insuperable obstacles to giving an adequate truth theory for natural languages (56, p.165). The problem is so well known that it hardly needs explaining. In a nutshell it is that under fairly innocuous conditions, which plainly seem to be satisfied by natural languages, the T-scheme, around which the theory of meaning is built, produces contradictions. And contradictions are, supposedly, intolerable, especially in the context of a logic which renders an inconsistent theory totally trivial.

The third problem concerns the connection between a theory of meaning and a speaker's understanding of his or her own language. For if we take it that knowledge of the content of T-sentences is not only sufficient for understanding language, but also necessary, it follows that a speaker knows propositions which involve essentially the notion of truth for that language. Now in general, to say that someone knows or believes a proposition is not to say that he or she understands a language in which that proposition can be expressed. Dogs can (it would seem) believe that a bone is buried at a certain spot in the garden. However, to have beliefs which are abstract and highly articulated structurally does seem to require the possession of a language in which they can be expressed. It follows that a speaker must speak a language which expresses the notion of truth for his/her language. Thus either we are off on an infinite regress, which will in fact get us nowhere, or the speaker's language is semantically closed and we are back with problem two: the semantic paradoxes. In a sense therefore, this problem is a simple but important corollary of the previous one.

Having briefly outlined the three problems, we will now consider each one in detail with an eye to the solution provided by a relevant bi-implication in the T-scheme.

4. The Weakness of Extensional T-sentences. The first problem is how to specify exactly what an adequate truth theory, and hence, what an adequate T-sentence, is. As we saw, the Tarskian way of doing this uses notions which are unavailable to someone who wishes to use truth theories to get at meaning.

The orthodox solution to the problem comes in three stages. First, given a putative truth theory we must, for every sentence s, designate one of the sentences of the form: s is true iff p, as the T-sentence for s.[10] Perhaps the best way of doing this is via the notion of canonical proof theory of the truth theory.[11] (The precise justification for selecting in this way is somewhat problematic. However, we will pass over this.) The idea is this. Given s, we determine the p in question as follows. Take s and work out what it is for an arbitrary sequence to satisfy it using the recursive satisfaction conditions in reverse (i.e. decomplexifying) order. We then eliminate the sequence, essentially by applying identities of the form ii) of section 2. (This can always be done since the formula is closed.) The result is the required p, and the computation procedure ensures that for this p, s is true iff p, is provable.

Having fixed on the set of T-sentences this way, the condition that they be provable (in Tarski's adequacy criterion for a theory of truth) now becomes vacuous. To compensate for this we must add as an additional requirement that the T-sentences in question be true (see e.g., Davidson 73, p.84). This is obviously a necessary condition for an adequate theory of truth. Unfortunately, however, it is not sufficient for a theory of truth which is to be the core of a theory of meaning - at least if the main connective of the T-scheme is a material equivalence. For then all that is necessary to meet this requirement is that the two sides of the bi-conditional have the same truth value. Thus

'snow is white' is true in English ≡ grass does not grow on cows

might do.

It might be thought that this unfortunate kind of situation would be ruled out by the formal requirement that the right-hand side of the T-scheme result from the use of the canonical proof procedure in the way described. It does not.[12] For example, suppose we have a truth theory for English such that the basis clause for the predicate 'x is a cordate' is

s satisfies 'x is a cordate' ≡ s($\underline{x}$) has kidneys,

or

s satisfies 's is a cordate' ≡ s($\underline{x}$) has a heart or grass grows on cows.

Then the canonical proof procedure will turn out true T-sentences which are obviously not sufficient for interpretative purposes. A discerning eye might note that the root of the problem here is exactly the material bi-conditional.

The orthodox proposal to avoid this problem is that the truth theory in question should not only produce true T-sentences, but that it should also survive "empirical testing".[13] The precise nature of this testing is usually a matter of some vagueness. However, the most plausible account is something like this: we test an individual T-sentence by seeing whether speakers assent to s when they may reasonably be expected to believe that p, command s when they may reasonably be expected to desire that p, etc.[14] A theory is better the more it maximizes the number of its T-sentences successfully tested. Thus, the ultimate testing of a theory is holistic.

Exactly how the empirical testing is supposed to rule out the deviant extensional T-sentences is never really discussed. Presumably, to take the first deviant example we gave, it is (in a rather idealised form) like this: we take a speaker to a dissection room and carefully dissect a novel species of animal to expose the kidneys. If the speaker, on the basis of this, will not assent to the claim that the creature is a cordate, the theory has a black mark.

A little thought shows how shaky this procedure is for ruling out deviant T-sentences. For it is quite possible that all speakers believe that all renates are cordates and will, in fact, respond 'yes' when asked whether, on the basis of this evidence, the creature is a cordate. This possibility is indeed heightened when we remember linguistic "division of labour". For someone competent to recognize kidneys *in vivo* is likely to know a good deal of biology already. It might be hoped that this kind of situation will be taken account of when we find out what other sentences the speaker will assent to, that is, that the holistic maximization will do the job. However, put like this it is clearly more of a hope than an argument. In fact it is most unlikely that this kind of procedure is guaranteed to rule out all deviant extensional T-sentences.[15] Suppose, for example, that we take for A some sentence which is obviously true, such as 'people normally have legs', then it is going to be virtually impossible to design a test which will elicit different responses to 'a is an X' and 'a is an X and A'. The whole idea of ruling out deviant extensional T-sentences must then be highly dubious.

The problem does not arise if we take for the main connective of T-sentences a relevant implication. For the deviant T-sentences are not then true. Indeed, the inferences from $A \longleftrightarrow B$ and $\neg C$ to $A \longleftrightarrow (B \vee C)$, and $A \longleftrightarrow B$ and C to $A \longleftrightarrow (B \wedge C)$ are invalid. Nor is there an entailment between having a heart and having kidneys. Thus, the proposed counterexamples no longer work. We may therefore simply revert to the requirement that T-sentences are true, for the purpose of characterising a truth theory adequate for a theory of meaning. To determine whether a proposed truth theory *is* true we may still have to resort to empirical testing. But such is the case with any sort of theory, whether in natural sciences, linguistics, or whatever. However, the point remains that going relevant removes the extensional spanners from the definitional works.[16]

5. The Semantic Paradoxes. Let us turn now to the second, and perhaps most crucial, reason for rejecting classical (or intuitionist) logic as the underlying logic of truth theory: the semantic paradoxes. The problem has been recognised as a central one for the Davidsonian position from its inception.[17] Unlike the first problem we discussed, however, little serious attempt has been made to face it.[18] There are several suggested solutions to the semantic paradoxes, and their several adequacies are highly moot.[19] Fortunately we need not go into the general question here. For the question is only what moves are available to a Davidsonian, given the other constraints on the enterprise.

Davidson's preferred solution to the problem is that no natural language is semantically closed, i.e. can express its own truth theory. As he puts it:

> There may in the nature of the case always be something we grasp in understanding the language of another (the concept of truth) that we cannot communicate to him (Davidson 67, p.314).

This is highly unsatisfactory; first, because what is at issue is not only the understanding of other people's language but also of our own language. Could it really be the case that when we understand our own language there is something we cannot communicate? We will return to this question in the next section. Secondly, the claim that a language is incapable of expressing its own truth theory is highly implausible. No one really believes it: if they did, those interested in the semantics of English would be rushing off to learn Urdu, Hindi, etc. in order to find the key to their problems. The central point is, of course, that to suppose that the truth theory of a language cannot be given in that language flies in the face of what Tarski called the universality of natural language (see 56, p. 164): that anything that can be communicated can be communicated in, e.g., English. Since Tarski wrote, some doubt may have been cast on this thesis by fashionable theories of incommensurability.[20] However, one cannot seriously mount a case that the notion of truth (in English) is incommensurable with the English vernacular (whatever, in the end, that is supposed to mean), precisely because that notion is part of that very vernacular. English speakers operate with this notion all the time. If, then, a semantics of English can be given, there is no reason (other than a desire to save an adopted position) for supposing that it cannot be given in English.[21] The English may be augmented by a certain amount of jargon and technicalities, but this is no more than is granted to any English-speaking scientist.

It should be noted that Davidson's view sits ill with another of his views (expressed a few years later) about the possibility of others having concepts different from ours. Davidson's view is not that it is impossible for others to have concepts we lack. It is the far more damning one that such talk has no real sense. Arguing (partly) on the basis of his methodological views on translation (which are essentially those for finding an adequate truth theory that we discussed in the last section) he concludes:

> It would be wrong to summarize by saying we have shown how communication is possible between people who have different schemes, a way that works without need of what there cannot be, namely a neutral ground, or a common coordinate system. For we have found no intelligible basis on which it can be said that schemes are different. It would be equally wrong to announce the glorious news that all mankind - all speakers of language, at least - share a common scheme and ontology. For if we cannot intelligibly say that schemes are different, neither can we intelligibly say that they are one (Davidson 73b, p.20).

If this is right, the very basis of saying intelligibly that someone possesses the concept of truth-in-English, whilst we do not, collapses.

What other lines are available to a Davidsonian on the semantic paradoxes? Given that the concept of truth (in-English) is representable in English, the other main possibility is that certain instances of the T-scheme fail. For these, plus classical (or intuitionist) logic and the referential potentialities of English give the contradiction.[22]

Now one might well argue that if the T-scheme is jettisoned, then truth isn't characterized. After all, that the T-scheme holds appears to be a minimal condition of adequacy on any characterization of truth. However, we will not pursue this here. For it is sufficient to point out that the move of rejecting the T-scheme is hardly open to a Davidsonian! For every sentence must participate in a true T-sentence. This, after all, is what gives its meaning. Thus, rejection of the T-scheme undercuts the very possibility of taking a theory of truth to be the basis of a theory of meaning.

It might be suggested that the fact that the T-sentence for a sentence, s, fails just shows that such a sentence is meaningless. Indeed, it is often mooted as a solution to the paradoxes that paradoxical sentences are meaningless. The suggestion has little value unless it can be made precise. The notion of meaningfulness required cannot be cashed out in any syntactic way.[23] However, the idea that the sentence has no true T-sentence in a theory of truth which is adequate as the basis of a theory of meaning, is a way of making this idea precise.

Of course, if we wish to take this line, we can no longer accept our previous account of what an adequate truth theory is. For it was part of this that every T-sentence is true. (See section 4.) Whether or not a new account can be given we do not know. Neither is it important. For this approach fails for Davidsonian purposes for quite separate reasons: such a theory of truth cannot be axiomatic.

Let us spell this out precisely. We have a language L, with a truth predicate T. A theory of truth for L in the language L is now no longer required to prove *all* T-sentences, but only those that are true (in some sense, the onus for producing which is on the proponent. However, an example might be truth at a fixed point in the Kripke Hierarchy (see Kripke 75)). Now suppose that such a truth theory is axiomatic (i.e. recursively enumerable). Then the set of T-sentences is axiomatic too (since we can effectively tell when we come to a T-sentence in the enumeration of theorems). Hence the set of sentences s for which the T-scheme is provable is r.e. too (since we can effectively determine s from its structural description). Call this set τ. The proof now requires three further assumptions:

1) L contains the language of first order arithmetic.

2) The sentences of first order arithmetic are unproblematic, i.e. are all straightforwardly true or false in the usual way. (So, technically, the standard model of arithmetic is a substructure of the interpretation of the language.)

3) A sentence is true or false iff the T-sentences for it holds.

1) is surely true of any natural language such as English.

2) holds on virtually all the mooted interpretations for a language with its own truth predicate. Moreover this is, perhaps, what ought to be the case, since it is only the appearance of the truth predicate itself in a sentence which is normally thought to pose problems.

3) needs further comment and we will return to it.

The proof can now proceed. Since τ is r.e. it can be defined by an arithmetic sentence of L with one free variable A(x) i.e. A($\underline{n}$) is true iff n is (the code of) a member of τ. By the usual diagonal procedure,[24] we can find a formula

$$\neg A(\underline{n}) \vee T\ \mathrm{Neg}(\underline{n})$$

whose code is n, where Neg is an arithmetic functor which represents negation on (codes of) formulas. Call this ϕ.

Now either $A(\underline{n})$ is true or it is false (i.e. $\neg A(\underline{n})$ is true), by 2).

If $A(\underline{n})$ is true then the T-scheme holds for ϕ (and $\neg\phi$).

Hence $T\underline{n} \equiv \neg A(\underline{n}) \vee T\ \mathrm{Neg}\ (\underline{n})$

so $T(\underline{n}) \equiv T\ \mathrm{Neg}\ (\underline{n})$. (α)

But ϕ is either true or false by 3), and ϕ iff $T(\underline{n})$ (by the T scheme) iff $T\ \mathrm{Neg}(\underline{n})$ (by α) iff $\neg\phi$ (by the T scheme). Contradiction.

If $\neg A(\underline{n})$ is true, then $\neg A(\underline{n}) \vee T\ \mathrm{Neg}(\underline{n})$ is true, i.e. ϕ is true. Hence the T-scheme holds for ϕ by 3), i.e. $A(\underline{n})$ is true. Contradiction.

The argument, which is but a version of the extended liar paradox, shows that the theory T is not axiomatic. In fact the same proof shows that it is not even arithmetic. Since axiomatizability is a key feature of a theory of truth suitable for the basis of a theory of meaning, this shows that this way of solving the paradoxes is not open to a Davidsonian of classical persuasion.[25]

The major possibility for escaping this conclusion lies in denying 3). One half of it we take to be unproblematic. If sentences are true/false, then they are certainly meaningful, i.e. (on the proposed model) the T-scheme holds for them. What might be doubted is the converse. Maybe there are sentences for which the T-scheme holds which are neither true nor false. However, this position runs into plenty of problems of its own. Suppose ϕ is such a sentence with code n. Then certainly ϕ is not true, i.e. $\neg T(\underline{n})$ holds. (We may suppose, after all, that we are dealing with a formalization of English.) But by the T-scheme $T(\underline{n}) \equiv \phi$ and hence

$$\neg T(\underline{n}) \equiv \neg\phi$$

By *modus ponens*, $\neg\phi$, i.e. ϕ is false. Contradiction.

The only other possibility for dealing with the semantic paradoxes that seems open to a Davidsonian is the rather desperate one of trying to outflank all these considerations by claiming that the truth predicate of English is not univocal, that English itself is a hierarchy of sublanguages, each of which contains the truth predicate for one of lower rank, the actual predicate of English doing typically ambiguous duty for all these. Implausible as this view may seem, it used not to be an uncommon view that English was a hierarchy of Tarski metalanguages[26] (despite the fact that this was never Tarski's view). Yet this view has now lapsed, and quite rightly. A natural language, such as English, can not be viewed in this way.

The arguments for this are well known and need no rehearsal here.[27] The crucial point is that there are perfectly good sentences of English which can not be fitted into this procrustean framework. Even a common or garden sentence such as 'Everything he said is true' may be such a sentence if empirical circumstances turn out unfavourably. And a sentence such as 'There is some sentence of English which is true in none of the hierarchy of languages which is [supposedly] English' could not be a sentence of any member of the hierarchy in any circumstances. We need not, therefore, discuss this view further.

Thus, as long as we stick to classical (or intuitionistic) logic, the semantic paradoxes seem to pose an insuperable problem for the theory of meaning with which we are concerned. However, the whole problem vanishes once we move to a relevant logic. For then we can simply allow the paradoxes in. The truth theory of English can be given in English. All instances of the T-scheme will be true, indeed, provable in the theory. Consequently, we will be able to use them to prove theorems of the form $\phi \longleftrightarrow \neg\phi$ and hence $\phi \wedge \neg\phi$, but that does not matter. For in a relevant theory contradictions can be localised; that is, they do not spread to the whole theory. The spread law of classical logic $(\phi \wedge \neg\phi) \rightarrow \psi$ fails.[28] Moreover, and particularly to the point, there is nothing specific about its application to a theory of meaning which requires a truth theory to be consistent. All that is required is that the theory specify a unique and true T-sentence for each sentence, s, of the language. And it can do this, consistent or not.[29]

Nor is there any technical problem here. In the construction of section 2, we simply let M be O. The existence of constants and function symbols in M is an inessential difference (as a simple extension of the construction along orthodox lines shows), and we can simply identify R^n_i with Q^n_i. We need one extra axiom for the truth theory since the object-language now has an extra predicate 'Sat'. This is the obvious:

$$\text{Sat s Pred}^2\ \underline{\text{Sat}}\ \text{xy} \longleftrightarrow \text{Sat } s_x\, s_y.$$

and with this extra axiom all instances of the T-scheme are provable in the usual way.

6. Speakers' Understanding. Let us turn to the third problem we will consider. In many ways this is but a corollary of the second problem. Still, it is a very important one, in fact crucial to meaning-theory. It is, however, a complex issue, where clear cut arguments are difficult to find. For this reason we will spell out the problem in the form of a sorites and then discuss each step separately.

α) A theory of meaning should make explicit the understanding or knowledge which constitutes a speaker's competence with a language.

β) Hence the speaker of a language must have propositional knowledge of the content of the T-sentences.

γ) Such knowledge requires that s/he should speak a language in which this is expressible.

δ) This is either a different language, in which case we are off on an infinite regress which is untenable,

ε) or else the language is the same, in which case it is semantically closed. But this option is closed off to us because of the paradoxes.

α) is a popular view of the function of the theory of meaning for a language. Obviously the notion of competence here must be an idealised notion; it is not intended that it must be instantiated precisely in any given speaker of the language. Yet on such a view of the role of a theory of meaning, the theory must be at least in principle knowable by any competent speaker; or, to be more precise, that portion of the theory which states explicitly the knowledge involved in such competence should be, in principle, knowable.

It must be pointed out that this view, that a theory of meaning must capture what a language speaker knows, is not mandatory. Indeed, Davidson himself seems not to endorse this requirement. For him, the condition that knowledge of the consequences of a theory of meaning should suffice for interpretation, functions only as an adequacy test ensuring that a theory *does* yield interpretative T-sentences. He does not hold that this must be what speakers actually use which enables them to interpret.[30] However, it is rather difficult to justify the interest in theories of meaning if all they are meant to do is to meet a requirement of sufficiency for interpretation. How then is it that they give information about language and meaning which those wishing to understand the nature and use of language have traditionally sought? If, to put it another way, this is not how speakers do interpret, then how do they? And once we have answered this question, the whole official "theory of meaning" becomes otiose. If the "cost" of a theory which shows how speakers *do* interpret is to change to a less crude account of the conditional, it would seem foolish to resist.

β) Given α) and an account of meaning of the kind we are discussing in this paper then β) follows. For it is precisely the content of the T-sentences which allows a speaker to interpret. Thus, for every sentence s, a speaker must know that s is true iff p, where p ... Now this does not in itself imply that the speaker must know any sentence of any language which expresses this fact. At least arguably non-language speakers such as animals can have propositional beliefs. What the speaker must know are the contents of, or propositions expressed by, the T-sentences. But ...

γ) to suppose that a speaker can have such knowledge when s/he has no language in which to express it leads us to a tangle of vexed problems centred around the question of what it could be to have such knowledge. Why, for example, would it be absurd to attribute such knowledge to a caterpillar; and what makes it any less absurd attributing it to a larger animal (such as a person) who can't say it? Whilst we can hardly hope to give satisfactory answers to such problems here, we will at least try to show that the idea that such belief could be "tacit" is faced with enough problems to make the alternative attractive.

A first step is to observe that before we are prepared to attribute "mental states" such as knowing, believing, desiring, etc. to animals, they must show a sufficient complexity of behaviour. It is only when an animal behaves in certain ways that we find it appropriate to explain its behaviour in terms of intensional states. Thus there seems little in the behaviour of a caterpillar which requires appealing to intensional states in order to explain it. By contrast, if a dog, upon hearing a noise at the door, barks, wags its tail, and generally appears in a state of anticipating a familiar person, it is reasonable to explain this by saying that the dog believes such a person to be at the door. Similarly, a dog who, without prior suggestion, brings its lead to its owner might reasonably have its behaviour explained by saying that it wishes to go for a walk.

Non-linguistic behaviour will, however, take us only so far in attributing intensional states. It seems unwarranted to attribute to the beleaded dog a desire to go for a three mile walk, a walk in Scotland, a walk with a person thinking about Kant, etc. Since we "have no good idea how to set about authenticating the existence of such attitudes when communication is not possible" (Davidson 74, p.312), they are quick to fall to the application of a razor. In general, therefore, it is necessary for a creature to demonstrate verbal behaviour and, in particular, make explicit reports of beliefs, intentions, etc. before we may reasonably attribute complex intensional states to it. So much, in general, even Davidson would agree with.[31] The crucial question now is this: what behaviour, linguistic or otherwise, is necessary for us to sensibly attribute knowledge of the propositions expressed by T-sentences for a language L; and in particular, is there any behaviour that will do this short of the person's ability to state these propositions in some language?

Perhaps the obvious suggestion is to take the person's ability to speak L as such behaviour. After all, this knowledge is precisely what is being invoked to explain the ability. However, a moment's thought shows that this is unsatisfactory. That someone has a certain knowledge has been proposed as an explanation of a person's ability to do certain things, viz. speak a language. If, when we ask what grounds we have for supposing a person to have such knowledge, all we can produce is the fact that the person speaks the language, we have gone round in a rather unilluminating circle. For that someone has such knowledge is, in the end, no more than that they speak the language. And this cannot, therefore, be offered as an explanation of their ability to speak it.

Another suggestion is that the behavioural evidence that shows that someone believes all instances of the T-scheme: s is true iff p, is that for any s the person is prepared to assert s when they believe that p, command s when they desire that p and so on. Yet this is nothing other than the ability to speak the language. It therefore gets us no further, as we have just seen.[32]

What is required is some purchase on the attribution of knowledge of the content of T-sentences independent of speaking the language. If we could ask the person certain questions about whether or not they accept certain T-sentences, this would provide such purchase.

However, this possibility is precisely ruled out by the supposition that the person speaks no language in which these can be expressed. We find it hard to see what other behavioural evidence there could be. Indeed, the following considerations suggest that there can be none: first, non-linguistic behaviour on its own can hardly be sufficient for the attribution of this knowledge. For if it were, it would be, in principle, possible for us to attribute understanding of language to a non-language user and this seems most implausible. But if the only additional behavioural evidence concerns the use of L, this too seems insufficient as we have seen. Thus this step of our sorites is highly plausible.

δ) Thus the knowledge must be expressible in some language understood by the speaker. Either this language is a public language or it is not. If it is a public language then this solves the problem only at the cost of posing it again at another level. Of course, the procedure can be iterated. However, no stopping point of finite level can be satisfactory. Hence it must be that the speaker speaks an infinite hierarchy of languages each of which can state the truth conditions of the prior one. Patently no one speaks an infinite number of languages. The regress is therefore vicious. The only hope here is the supposition that each natural language is in fact an infinite hierarchy of Tarski-metalanguages. We discussed this possibility in the last section and saw that it would not do.

So if the language in which the truth conditions are expressed is not a public language, what else can it be? Private languages are justly out of favour (since Wittgenstein 53). There is a suggestion by Fodor (75, ch.2) that speakers may, indeed must, have a non-public language and that, moreover, if this is conceived by analogy with the machine language of a computer (as opposed to a high level language) then Wittgensteinian problems are sidestepped. The correctness of Fodor's view raises many interesting problems, but we will not discuss them here. For it is unlikely, even if Fodor is right, that this will help with the problem at hand. For since the speaker is a fluent user of this language, we can ask how it is that the speaker understands what s/he is doing.[33] Thus we are off on a regress again. In fact the suggestion that the second language is private has added nothing to the argument.[34] Hence we arrive at ...

ε) The language which expressed the truth conditions must therefore be the language L itself, i.e. L must be semantically closed; but this option is ruled out as long as we stick to classical (or intuitionist) logic. The substantive steps here have already been covered in the last section and we need not repeat them.

So much for the problem. Its relevant/paraconsistent solution will be clear from the previous section. There are no insuperable problems regarding the idea that a natural language can state its own truth conditions, once we drop the idea that contradictions are trivialising. It then becomes quite clear that we can allow that the understanding a speaker has of his/her language is the knowledge of the content of certain T-sentences, and that this knowledge is expressible in that language itself. The problem (as conceived by a Davidsonian) of how a speaker understands his/her own language is therefore solved.

7. Conclusion: Consistency and Extensionality. The last two problems we have discussed are solved by allowing the truth theory for a theory of meaning to be inconsistent, but (we hope) non-trivial - a possibility allowed for by relevant logic, though not by classical logic. There will be some who will be less than happy about it. The very idea that we might be able to make good use of an inconsistent theory will be anathema. But once inconsistency no longer leads to triviality, what is there against this? The main argument would, presumably, be something like this: given any theory in science, linguistics, mathematics or elsewhere, it cannot be accepted unless it is true (or at least, can only be accepted in the knowledge that it is at best an approximation which needs to be replaced). Since no contradictions are true, no inconsistent theory, including the theory of truth we have specified, is acceptable.

What is to be said about this? One possible line is simply to adopt an instrumentalist attitude towards theories, that is, to take them to be mere combinatorial calculating devices which may churn out the right predictions, but for which the question of truth does not matter, or does not even arise. In this case, the truth "theory" would be a sort of black box which turned out pairs of sentences, the first mentioned and the second used - the pairs correlated by the T-sentences. However, this line is not acceptable as it stands. There are general philosophical problems concerning instrumentalism to be faced.[35] But more pertinently here, it was a criterion of adequacy on any theory of truth that its T-sentences be *true* (see section 4). We might try to get round this problem by allowing the T-sentences to be true/false, and relegating the rest of the "theory" to a merely instrumental role. However, it is not clear that this will solve the problem. For the inconsistent part of the theory is almost certain to spread into the T-sentences themselves. Suppose the referential machinery of the theory is sufficient to give us a liar-sentence, i.e. a sentence of the form 'a is not true', whose name is 'a'. This will happen if the language in question is a natural one such as English. Then we can prove its T-sentence 'a is true iff a is not true', and hence both 'a is true' and 'a is not true'. Now provided the underlying logic contains the rule of inference $A,B/\neg(A \rightarrow \neg B)$,[36] or even the special case $A/\neg(A \rightarrow \neg A)$, we can then infer 'it is not the case that (a is true iff a is not true)' i.e. the negation of the T-sentence. Hence the contradictions have spread into the non-instrumental part of the theory.

The fact that it seems unlikely that the contradictions can be kept out of the T-sentences suggests that we take a very different line of reply to the initial objection. We can simply deny that the fact that the theory has contradictions in it prevents it from being true. Heretical as this may sound, there is mounting evidence that this is the right line. First, the very semantics of relevant logic allow the possibility that contradictions are true - in at least some "possible worlds",[37] and it is difficult to see what it is about this one which prevents it from being one of these. There is, at any rate, no problem about a formal semantics which allows for true contradictory theories. Secondly, recent discussions of the logical paradoxes suggest that the correct way of treating the paradoxical sentences, is precisely as true contradictions.[38] (So the actual world *is* inconsistent.) This fact is doubly relevant in the present context since the inconsistencies in the theory of truth just are the semantic paradoxes and their derivatives. Third, it used to be a common view that the notion of truth in a natural language such as English is inconsistent.[39] Exactly what this amounted to was never, perhaps, spelt out in detail. However, because of the influence of classical logic, inconsistency was taken to imply incoherence, and it was therefore thought that the natural language conception had to be ditched. On the present view, the claim that the notion of truth in English is inconsistent is reinstated. This can be spelt out precisely. To say that the notion is inconsistent is to say that the principles governing the notion (i.e. the axioms of the truth theory) plus, possibly, some empirical facts about reference determine the truth of some sentences of the form $A \wedge \neg A$. It should also be noted however, that in virtue of this, the T-scheme and the standard truth conditions for conjunction, '$\underline{A}$ is true and $\underline{\neg A}$ is true' also follows. Thus the notion of truth not only produces contradictions but, so to speak, allows for them. This conception of truth, if inconsistent, is not incoherent. Only a lapse into triviality or near triviality would destroy coherence. Nor, as we have already observed, is there anything which prevents the T-sentences of such a theory (even those which are true and false) from being the meaning-giving parts of a theory of meaning.[40]

It is certainly possible to raise other objections to the idea that there are true contradictions. (Their cogency is, however, another matter.) But these are not specific to the present proposal concerning truth and meaning. Hence this is not an appropriate place to discuss them.[41]

We have seen that adopting a relevant/paraconsistent theory of truth is a live option for a Davidsonian approach to the theory of meaning, and, moreover, that doing so resolves a number of pressing problems. It would be surprising if human ingenuity were not able to suggest solutions to the problems which allowed the retention of classical logic. These must, of course, be treated on their merits. Yet it is reasonably clear from the discussion that such "classical" problem-solutions are likely to be complex and have an air of artifice about them. Further, it is unlikely that classically the problems can be solved uniformly, as they are by adoption of a relevant logic. The weight of evidence therefore seems to be firmly in favour of relevance. Against this, conservatism is bound to assert itself. One bulwark of this conservatism is extensionalism: the view that non-extensional functors such as the $\rightarrow$ of relevant logic are of dubious intelligibility. However, Davidsonians, far from being trapped with this view, should be emancipated from it. For the $\rightarrow$ of relevant logic is shown to be intelligible and meaningful by the very account of meaning which they endorse: in section 2 of the paper we saw that a theory of truth could be given for a language, O, with a relevant implication $\Rightarrow$. Of course, this was done by using a relevant implication in the metalanguage. But this, *per se*, is no objection, or the intelligibility of the extensional connectives would itself fall to it. Thus, if arguments for extensionalism are to be found, these must come from outside the Davidsonian programme. And, as we have seen, these can cause it nothing but trouble.

NOTES

The idea is Davidson's. See, e.g., Davidson 67 or 75. It has been developed by others. See, e.g., McDowell 76, Platts 79 Ch.2., Davies 81.

It has also been suggested that the set of axioms should be not merely decidable but finite. However, the arguments for this seem to us not very cogent.

The relevant truth theory we go on to discuss uses an intensional functor, $\rightarrow$, but does not quantify over intensional entities. Hence it does not renege on the original Davidsonian heuristic.

See, e.g., Anderson and Belnap 75, Routley *et al.* 82.

The set of individual constants of M could be finitized if we were to take the variables and predicates of O to be generated from a finite vocabulary.

Unfortunately, traditions have it that 's' is used for both evaluations (sequences) and sentences. In this section we follow the former tradition; in all others, the latter.

It is not difficult to replace the scheme i) by a finite set of axioms provided the variables of O are generated as in fn.5.

It would be more accurate to say that the T-sentences provide the interpretational content for sentences, to which we may then need to apply further rules amounting to a theory of force in order to fully interpret utterances of those sentences. But this aspect of the theory need not concern us here.

Strictly speaking, knowledge of the T-sentences is not quite sufficient to be able to interpret a language. One must also know that the T-sentences *are* meaning-giving, i.e. that they are "entailed by some true theory that meets the formal and empirical constraints" (Davidson 73a, p.327). Again, this ramification need not concern us here.

This step is absolutely necessary. It is unfortunate that many writers on the subject have appeared to fail to notice this. (See Priest 80a.) Writers discussing stages two and three frequently refer to *the* T-sentence for s without noticing that its uniqueness depends upon the notion of translation.

Suggested to us by Barry Taylor. See also Davies 81 pp.33-4.

As is pointed out, e.g., by Foster 76 and Loar 76.

E.g., Davidson 76, p.321-2:

> There is no difficulty in rephrasing Convention T without appeal to the concept of translation: an acceptable theory of truth must entail, for every sentence s of the object language, a sentence of the form: s is true if and only if p, where "p" is replaced by any sentence that is true if and only if s is. Given this formulation, the theory is tested by evidence that T-sentences are simply true; we have given up the idea that we must also tell whether what replaces "p" translates s. It might seem that there is no chance that if we demand so little of T-sentences, a theory of interpretation will emerge. And of course this

would be so if we took the T-sentences in isolation. But the hope is that by putting appropriate formal and empirical restrictions on the theory as a whole, individual T-sentences will in fact serve to yield interpretations.

This account is closest to McDowell's. See Platts 79, Ch. 2. The *exact* details do not seem to matter too much since any proposal along these lines seems to fall to the kind of objection we shall bring.

Indeed, as is well known, Quine has argued that this kind of process cannot even rule out extensionally incorrect T-sentences, on the grounds of the flexibility of global adjustments. See, e.g., his account of the indeterminacy of translation in Quine 60, Ch.2.

There remains the problem of truth theories which produce intensionally equivalent variants of the meaning-giving T-sentences. E.g., s is true iff $\neg\neg p$ (or iff $p \wedge p$ or iff $p \vee p$ or iff $p \vee (p \wedge q)$). (Notice that this was a problem for the Davidsonian approach anyway. No empirical testing procedure, it would seem, can get rid of such intensionally deviant T-sentences, precisely because the right hand sides are logically equivalent.)

There seem to be two possible lines here. One is to say that all logical equivalents have the same meaning, and hence that the intensional variants are not counterexamples to the 'T-sentences give meaning' thesis. This line is very difficult to maintain if logical equivalence is classical logical equivalence. However, relevant logical equivalence is a much tighter notion, and with logical equivalence so understood, this line is a difficult one to refute. The fact that two things mean the same does not, of course, imply that anyone necessarily realises they do. The other possible line is to add a further condition on what counts as an acceptable T-sentence, to the effect that the logical structure of p must be as simple as possible commensurate with the logical structure shown syntactically by s. (Notice that such a constraint will not solve the problem for the classical Davidsonian of the failure of the empirical constraints, as can be seen by considering the heart/kidneys example.) The precise formulation and rationale of this constraint clearly need more investigation.

See Davidson 67, pp.313-15.

To an extent this may be due to the fact that according to Davidson the matter may be set aside for the moment. "The first point [viz. the semantic paradoxes] deserves a serious answer, and I wish I had one. As it is, I will say only why I think we are justified in carrying on without having disinfected this particular source of conceptual anxiety". Davidson 67, p.314. As we shall see, the problem cannot be set aside this easily. It is not a peripheral problem but strikes at the very root of the enterprise.

See, e.g., Priest 84a.

See, e.g., Feyerabend 75, Ch.10.

That is, the claim is obvious to anyone not in the grip of a philosophical theory.

See Tarski 56, section 1, or Priest 84a.

As, e.g. Kripke has argued. See Kripke 75.

See, e.g. Boolos and Jeffrey 74, p.176.

It should be pointed out that since the above argument uses the disjunctive syllogism, it is not one that a relevant logician would, in fact, accept as correct. However, the argument is *ad hominen* against a classical (or intuitionist) logician and, as such, is perfectly in order.

Something like this view was certainly held by Russell at one time. See Russell and Whitehead 10, p.41ff.

See Kripke 75, Gupta 82, Priest 84a.

There is an important point here, however. The relevant logic used must be one in which the absorption principle $A \rightarrow (A \rightarrow B)/(A \rightarrow B)$ fails. If one of the elect Anderson-Belnap logics (*E*, *R* or *T*) is used, triviality does result due to Curry paradoxes. See Priest 80. Strictly speaking, what is necessary to show that contradictions do not spread everywhere, is a non-triviality proof. Though we are fairly sure that the semantically closed theory we go on to give is non-trivial, we have, as yet, no proof of this.

That a suitable relevant logic permits a paraconsistent approach to the paradoxes has been repeatedly pointed out in the literature, e.g. in Priest and Routley 83. What concerns us here is precisely the importance of this for the theory of meaning.

See, e.g., Davidson 73a, p.313.

"Making detailed sense of a person's intentions and beliefs cannot be independent of making sense of his utterances". Davidson 74, p.312. Indeed, Davidson holds the much stronger belief that we cannot attribute *any* intensional states to a non-speaker. See Davidson 73a.

The suggestion also runs up against problems posed by the fact that many sentences have truth conditions which transcend recognition. See e.g., Dummett 76, pp.79ff.

"I am committed to the claim that [the language of thought] is, in a certain sense, understood: e.g., that it is available for use as the vehicle of cognitive processes." Fodor 75, p.65.

Fodor does suggest (75, p.66) that what constitutes a machine's "understanding" of its machine language is an isomorphism between the structure of the language and the machine states. The same is meant to apply to the language of thought, with, presumably, 'machine states' replaced by 'brain states'. It might be thought that this provides a way of breaking out of the regress. However, what this possibility does (if indeed one can ultimately make sense of it) is something much more radical. For the correlations between public language and language of thought, and language of thought and brain states induce one between public language and brain states. And if the language of thought/brain states correlation is an adequate account of what constitutes the understanding of this language, then, presumably, the induced correlation gives an adequate account of what constitutes the understanding of the public language. The language of thought drops out as an unnecessary intermediary. We would then have a reductionist account of understanding and hence of meaning for natural language. Far from solving a problem for the Davidsonian programme, it would therefore undercut the whole thing. Some of these points are discussed further in Crosthwaite 1983, (Chapter 7).

See e.g., Smart 63, Ch.2.

It should be noted that many relevant logics do not have this principle. However, it is certainly possible to have a relevant logic containing it which is not Curry-trivialised.

See Priest 80 or, for a genuine possible world semantics for relevant logic, Routley *et al.* 82.

See Priest 79, 84a, 84b and Priest and Routley 83, Ch.5 section 2.

For references, see Herzberger 67. For further discussion of the following points, see Priest 84a.

It might be suggested that the fact that 's is true iff p' and 'it is not the case that (s is true iff p)' are both provable in the theory shows that according to this theory s both means and does not mean p; and that this is objectionable. However, this is false on both counts. First one need not suppose that 's is true iff p' is synonymous with 's means that p'. 's means that p' is to be cashed out in terms of the appropriate T-sentence being provable in a certain truth theory. But even if this were not the case, what is there against a theory of meaning having contradictory singularities?

A discussion can be found in Priest and Routley 83, especially Ch.5 section 3.