

6

Theories of Meaning
and Speakers' Knowledge

This paper is concerned with recent – 1970s – trends and emphases in British philosophy of language, particularly as practised in Oxford. The question, what form should be assumed by a satisfactory theory of meaning for a natural language, quite suddenly came to seem absolutely fundamental, with the most comprehensive philosophical insights at stake. My concern here, however, is not to survey that issue, or to comment on the general significance it was (and still is) widely believed to have, but to explore certain assumptions which those who led opposite sides in the debate – principally Professors Davidson and Dummett – seem to have had in common: assumptions whose correctness is arguably not just essential for the philosophical significance – even a fairly local philosophical significance – of a 'theory of meaning' but is tacitly presupposed by the greater part of all philosophical semantic endeavour.

I

Davidson's proposal¹ that a theory of meaning for a natural language should take the form of a recursive truth definition in the style which Tarski devised for certain formal languages is opposed, or qualified, by Dummett in three principal respects. First, such a theory could be at best the core of a fully fledged theory of meaning; it would need to be embedded within a theory of *force*, a theory concerned with the difference between assertion, command, question, wish and other modes of illocution, and which made plain how these different types of speech act were signalled in the language

1 The *locus classicus* is his 'Truth and Meaning'. The paper is reprinted as Essay 2 in his *Inquiries into Truth and Interpretation*. His book, and especially the section consisting of the first five essays, is pervaded by the proposal.

under study.² Second, whereas Davidson seems content, in the case where the language in which the theory is formulated includes as a proper part the language theorized about, that the theory should be *homophonic* – that is, that the form taken by its meaning-delivering theorems should involve *use* of the very sentences whose meaning is being characterized – Dummett requires that the theory should be *full-blooded*, that it should, where possible, analyse and illuminate meanings.³ Third, Dummett has disputed Davidson's choice of truth as the central notion in terms of which an account of meaning should be given: if what the theory says about any particular sentence aspires to be what anyone familiar with the meaning of that sentence knows, there are 'anti-realist' arguments, made familiar by Dummett, why it cannot in general suffice for the theory to assign truth-conditions to sentences.⁴

These are substantial differences. But they all concern what Dummett sees as deficiencies in the *output* of a theory of the projected kind – the first two being deficiencies of omission, while the third, though absolutely fundamental at the level of interpretation, may require no substantial alteration of the form of a Davidsonian theory. Dummett's criticism does not express any antagonism to Davidson's overall project; indeed, the points of agreement seem more striking than the disagreements. Davidson intends that a theory of truth is acceptable as a theory of meaning only if

- i it is finitely axiomatizable and
- ii it delivers a T-theorem for each declarative sentence of the object language in a manner that reflects the semantic structure discerned in that sentence.⁵

Why should the theory be finitely axiomatizable? Different passages in Davidson's writings suggest slightly different answers. Sometimes the thought is that understanding a language involves the capacity to make

2 See e.g. M. Dummett, 'What is a Theory of Meaning? (II)', pp. 72–6 and *passim*.

3 See M. Dummett, 'What is a Theory of Meaning?', pp. 102–6 and *passim*.

4 These arguments are widespread in Dummett's writings. See especially 'What is a Theory of Meaning? (II)' and Essays 1, 10, 11, 14 and 21 in his *Truth and Other Enigmas*. Note, however, page xxii of the latter's Preface where Dummett expresses dissatisfaction with the idea that the anti-realist arguments must call in question the propriety of truth-conditional theories of meaning, suggesting instead that their bearing is on the proper interpretation of truth. Cf. Essay 9 of this volume.

5 See the indexed references under *finiteness requirement* in *Inquiries into Truth and Interpretation*, and pp. 56–7 (of Essay 4, 'Semantics for Natural Languages') which the indexer missed. For those quite innocent of these matters, the 'T-theorems' take the form illustrated by

'Snow is white' is true if and only if snow is white.

whereby the quoted sentence is used to characterize its own truth-conditions and hence – controversially – its own meaning.

sense of no end of distinct expressions, and that this (potential) infinity contrasts sharply with the finitude of our capacities in general (most relevantly, presumably, the finitude of our capacities for information storage). A finitely based theory of meaning of the sort he is recommending will, Davidson believes, give us an insight into

...how an infinite aptitude can be encompassed by finite accomplishments.⁶

A neighbouring but distinct point is that if a language admits of characterization by a finitely based Davidsonian theory, then we have an insight into how the language can be learnable.⁷ The point is distinct because creatures with infinite capacities for information storage but no other infinitary abilities could only learn finitely much in a finite time. Finally, not quite the same thought as either of these is the idea that:

Speakers of a language can effectively determine the meaning or meanings of an arbitrary expression (if it has a meaning) and...it is the central task of a theory of meaning to show how this is possible.⁸

The theory of meaning is thus to contribute towards the explanation of how speakers can understand sentences which are novel to them. This thought is different because the question would arise – if it is a good question at all – for infinite creatures too, and for finite languages which, while semantically structured, possess no indefinitely iterable devices of the sort which generate a potential infinity of meaningful sentences.

However, the marginal distinctions among the questions, 'How can finite minds have infinite abilities?', 'How can languages be learnable?', 'How can speakers determine the meanings of novel utterances?', are not important. What is important is that the capacity of Davidsonian theory to assist in the provision of answers to any of them requires that it be admissible to think of actual speakers as equipped with the information codified in the axioms of a successful Davidsonian theory, and as prone to deploy that information in ways reflected by the derivations of meaning-delivering theorems afforded by the theory. Whether the theory aspires to cast light on *our* ability, finite as we are, to master a potentially infinite language, or *our* ability to complete the learning of the language, or on *our* ability to understand novel utterances, or on all three, success must depend, it seems, on its being permissible to suppose that it encodes information which *we* actually possess.

If Davidson is somewhat inexplicit about this, Dummett is not. A very definite commitment to the idea that the explanatory ambitions of a theory

of meaning depend upon recourse to some idea of speakers' *implicit knowledge* of its axiomatic contents is evinced in his writings on the topic. For instance:

A theory of meaning will, then, represent the practical ability possessed by a speaker as consisting in his grasp of a set of propositions; since the speaker derives his understanding of a sentence from the meanings of its component words, these propositions will most naturally form a deductively connected system. The knowledge of these propositions that is attributed to a speaker can only be an implicit knowledge. In general, it cannot be demanded of someone who has any given practical ability that he have more than an implicit knowledge of those propositions by means of which we give a theoretical representation of that ability.⁹

In an earlier paper, Dummett refers to

...our intuitive conviction that a speaker derives his understanding of a sentence from his understanding of the words composing it and the way they are put together.¹⁰

and relates Davidsonian theory to this intuitive conviction by the remark that:

What plays the role, within a theory of meaning of Davidson's kind, of a grasp of the meanings of the words is a knowledge of the axioms governing these words.¹¹

Further:

It is one of the merits of a theory of meaning which represents mastery of a language as the knowledge not of isolated, but of deductively interconnected propositions, that it makes due acknowledgement of the undoubted fact that a process of derivation of some kind is involved in the understanding of a sentence.¹²

For Dummett, the explanatory ambitions of a theory of meaning would seem to be entirely dependent upon the permissibility of thinking of speakers of its object language as knowing the propositions which its axioms codify and of their deriving their understanding of (novel) sentences in a manner mirrored by the derivation, in the theory, of the appropriate theorems.

There is accordingly a case for saying that, whatever their differences, Dummett and Davidson are in broad agreement about the interest of the project of a theory of meaning of this sort, and about the manner in which such a theory needs to be interpreted if it is to sustain that interest; and that they share a broad, underlying assumption about the nature of

6 *Inquiries into Truth and Interpretation*, Essay 1 ('Theories of Meaning and Learnable Languages'), p. 8.

7 See e.g. 'Truth and Meaning', opening paragraph.

8 *Inquiries into Truth and Interpretation*, Essay 2 ('Truth and Meaning'), p. 35.

9 'What is a Theory of Meaning? (II)', p. 70.

10 'What is a Theory of Meaning?', p. 109.

11 *Ibid.*

12 *Ibid.*, p. 112.

linguistic competence, viz, that it is fruitfully to be compared – at least in its basics – to any open-ended computational ability which – like, say, the ability to do simple arithmetical multiplications – deploys finite information in rule-prescribed ways. The difference is just that the knowledge which constitutes understanding of a language is gained, for the most part, not from an explicit statement – in contrast, e.g. with the multiplication tables, or the rules of chess – but by immersion in the practice of speaking the language in question.

That we understand novel utterances because we understand the words in them and significance of the way in which they are put together is apt to strike one as a platitude. But it is no platitude that the sort of project which Dummett, explicitly, and Davidson, implicitly, seem to have in view makes philosophical sense. If the gap is not immediately apparent, it ought to suffice to reflect that the platitude need only be regarded as describing a feature of the 'grammar' of *misunderstanding*: nobody may properly be described as misunderstanding a sentence unless guilty of some more specific misunderstanding, either of words deployed within it or of its syntax. Contraposing, we have the incontestable claim that if someone understands the vocabulary and syntax of a significant sentence, then they understand the (type) sentence. Platitude is left behind when the antecedent of this conditional is taken to describe an ulterior state of information which *enables* a subject to understand the sentence.

This paper is concerned with a number of interconnected questions. Let us say that a theory is *compositional* just in case it meets the two Davidsonian constraints, (i) and (ii), noted above. Section II is concerned with the question whether there can be any good motive for insisting on compositionality in theories of meaning which *avoids* recourse to the idea of actual speakers' implicit knowledge. Baker and Hacker¹³ have recently described Davidson's comparative inexplicitness on the topic as 'artful' – their thought being, I imagine, that, for the sort of reason outlined above, the questions on which Davidson hopes for illumination cannot be answered by Davidsonian theorizing unless implicit knowledge is invoked, but that Davidson has preferred not to elaborate on the issue, perceiving it for the Pandora's Box which they believe it to be. However that may be, it is important to be clear whether, besides the three mentioned, there is any different question, not directly concerned with the capacities of actual speakers, which devising a compositional theory of meaning might help to answer. Only if there is not is there any danger, should it emerge that there is absolutely nothing to be made of the notion of implicit knowledge, that Davidson's project will prove to have been a waste of time. Section III then takes up the question, what, if any, defensible conception of

implicit knowledge may be attained? Is there cause for confidence that a notion can be made good which is apt to allow a compositional theory to be explanatory of actual speakers' capacities?

If no suitable notion of implicit knowledge can be extricated, it naturally occurs to one to wonder how that leaves matters with the realism/anti-realism dispute in the theory of meaning. For the essence of the most influential anti-realist critique of classical semantics is exactly that it cannot be a theory of speakers' understanding: that it places impossible demands on the range of what speakers may reasonably be regarded as (implicitly) knowing. The anti-realist charge is that if 'truth' is understood as the realist intends, there is, in an important class of cases, no such thing as knowing the truth-conditions of a sentence. If this is to be a criticism of realist semantics, it is essential that understanding may be regarded as a species of knowledge. And it is obvious that if it may, it cannot *everywhere* consist in *explicit* knowledge. The ability to paraphrase a sentence is a (defeasible) ground for crediting somebody with understanding it, but it is not an ability which we have for a large class of sentences which we think we understand, (and our language might, in any case, have been such as to contain the resources for no such paraphrase.) Accordingly, if understanding is knowledge at all, the kind of knowledge it *essentially* is must, it seems, be implicit. An important argument running through Dummett's writings¹⁴ is exactly that only an anti-realist theory of meaning can sustain the demands made on the notion of implicit knowledge which, as we have seen, he believes that the theory of meaning must make. Section IV will contend that the anti-realist critique is not undermined by the worries about the notion of implicit knowledge whose well-foundedness there is most cause to suspect.

II

Doubts about the notion of implicit knowledge have surfaced frequently in the literature, even in the writings of those sympathetic to the spirit of the Davidsonian project. John Foster, for instance, writes

...having seen the generality of the theory required, we may wonder whether we should ascribe it to the speaker at all. The knowledge we would have to attribute to him is not, typically, what he would attribute to himself. His mastery of English equips him to interpret its expressions, but not to state the general principles to which these interpretations conform. Is it not

13 Baker and Hacker, *Language, Sense and Nonsense*, p. 324.

14 See footnote 4 above.

unnatural, even incoherent, to ascribe states of knowledge to which the subject himself has no conscious access;¹⁵

But Foster believes that the issue can be side-stepped:

... we can capture all that matters to the philosophy of meaning by putting the original project the other way round. Rather than ask for a statement of the knowledge implicit in linguistic competence, let us ask for a statement of a theory whose knowledge would suffice for such competence. Instead of demanding a statement of those metalinguistic facts which the mastery of a language implicitly recognises, let us demand a statement of those facts explicit recognition of which gives mastery. What we are then demanding is still a theory of meaning, but without the questionable assumption that one who has mastered the language has, at some deep level, absorbed the information which it supplies. The theory reveals the semantic machinery which competence works, but leaves undetermined the psychological form in which competence exists.¹⁶

There are two thoughts here. First, there is the idea that the quest for a theory of meaning does not need to be motivated by the desire to understand the capacities of actual speakers of a given language. It is enough, in Foster's view, that we seek to describe knowledge which *would* generate those capacities, whether or not it is the source of actual speakers' possession of them. If it is wondered why we should seek to do that, Foster's answer seems to be, second, that what the theorist of meaning is interested in is primarily *the way the language works*. While we should not, perhaps despair of the possibility that the theory might illuminate the 'psychological form' of actual speakers' linguistic competence, the primary object is to describe the 'semantic machinery' which drives the language.

The trouble with this is that the demands which Foster is making of the notion of meaning – the demands implicit in the image of 'semantic machinery' – threaten no less conceptual strain than the demands on the notion of knowledge – those generated by its qualification as 'implicit' – which he is trying to avoid. The intuitive response to Foster's proposal would be that it generates an intolerable divide between the concepts of meaning and understanding: truths about meaning have to be, ultimately, constituted by facts about understanding, so to aspire to a theory which aims to describe 'semantic machinery' independently of any assumption about what speakers of the language know is to aspire to a theory with no proper subject matter.¹⁷

¹⁵ From his 'Meaning and Truth Theory', pp. 1–2.

¹⁶ *Ibid.*, p. 2.

¹⁷ Compare the 'metaphysical perspective' of Elizabeth Fricker's, 'Semantic Structure and Speakers' Understanding', section I.

This response may involve over-simplification – depending on how 'ultimately' is understood – but it has great force, it seems to me, at the level of semantic primitives: expressions whose meanings are independent of the meanings of all other expressions of the language except those of which they are constituents, and to which a Davidsonian theory would devote its proper axioms. I do not think we can attach any content to the supposition that such an expression has a meaning except in so far as meaning is thought of as constituted, at least in part, by *convention*; and I do not think we can attain an account of the distinction between a convention and a corresponding regularity except by invoking the idea of practitioners' *intention*, qualified in various ways, to uphold that regularity. If both these, admittedly very vague, thoughts are correct, then the proper standing of the axioms of a theory of meaning must, it would seem, be grounded in speakers' intentions; and Foster's apparent belief that the theory can have an autonomous subject matter is of doubtful coherence.

There is another objection, perhaps less fundamental but more immediately clinically fatal. The fact is that there seems to be no necessary connection between Foster's recommendation that the theory should describe information which *would* suffice for mastery of a given, typical natural language and the constraints of compositionality. Or better: there seems to be no such connection in the case on which Davidsonians typically concentrate – the homophonic case. If one is aiming at construction of a heterophonic theory – whether because the object-language is quite different from that in which the theory is to be couched, or because one is aiming at a high degree of *full-bloodedness* in Dummett's sense – it may very well be that there is no way of completing Foster's task unless one aims for compositionality; that no other approach can effectively provide meaning-delivering theorems for every declarative sentence in the object-language. In the homophonic case matters stand differently. Provided we have a recursive specification of the syntax of the (declarative part of the) language, and provided we are content with the disquotational form of meaning-delivering theorem for which theories of truth are famous, Foster's project is well enough served by a semantic 'theory' which merely stipulates as an axiom every instance of the schema:

A is T if and only if P,

where 'P' may be replaced by any declarative sentence of the object language and 'A' by the quotational name of that sentence. This theory is not finitely axiomatized, but it is finitely *stated* and, in conjunction with the appropriate recursive syntax, it does yield the means for effectively arriving at a meaning-delivering theorem – assuming we have no independent reservations about truth theories on that score – for each

declarative sentence in the object-language. It thus fits Foster's bill: it describes information whose possession would suffice for mastery of the (declarative part of the) object-language. At least, it does so provided a compositional Davidsonian theory whose T-theorematic output coincided with the axioms of this theory would do so. It is true, of course, that the non-compositional theory could not be used to *impart* this information to someone who did not already have it – but then no homophonic theory, whether compositional or not, fairs better in that regard. The moral is simple: the ambition to describe information which would suffice for mastery of a particular language may impose certain constraints on the form taken by the theorems of a theory of meaning, but it imposes no interesting constraints on the mechanics of the theory.

We are looking for a project whose execution would call for a compositional theory of meaning but which would not demand that actual speakers be deemed to know the full contents of that theory. Foster's thought, in effect, was to idealize the language with a view to a theory of how *it*, autonomously, works. A quite different thought would aim to see compositionality as called for by the ambition to describe not some body of knowledge which speakers putatively have but what they are typically able to *do*. Suppose we essay to regard each of the T-theorems of a Davidsonian theory as descriptive of a sub-competence which someone who fully understands the object language has: a sub-competence constituted by sensitivity, in using the relevant sentence, to the constraint which the appropriate T-theorem captures. Have we completely described the general competence in which mastery of the (declarative part of the) relevant language consists when we have a theory which correctly describes all these sub-competences? There is a *prima facie* persuasive reason why we should ask for more. No matter what ability we are concerned to describe, and however complete our characterization of its ingredient abilities, the description is incomplete if the ingredients have certain causal interrelations about which it keeps silence. Someone who has a strong tennis game may have a good drive, a good lob, and a good slice on his backhand wing, but a *full* description of his skills would not restrict itself to the statement that each of these strokes is dependable if, let us suppose, the lob and the drive – when, unusually, either is fragile – tend to break down together, although the slice tends to remain a strong shot even when the rest of his game is off-colour. So with the theory of meaning: a full description of the competence possessed by speakers should not merely characterize its ingredients but ought also to reflect their (causal inter-)relations. This may inspire what Martin Davies¹⁸ calls the *Mirror Constraint*. Suppose it is true of speakers (a) that once they know what

18 Martin Davies, *Meaning, Quantification, Necessity*, chapter III, p. 53 and following.

S_1, \dots, S_n mean, they are able to know what the distinct sentence, S , means without any further exposure to the use of the language; and (b) that if induced to revise their belief about what S means, they would need no further inducement to revise their beliefs about what some of S_1, \dots, S_n mean. Then the Mirror Constraint says simply that if, and only if, speakers' sub-competences with S, S_1, \dots, S_n are so interrelated, an adequate theory of meaning for their language should ensure that those of its resources which suffice for the derivation of meaning-delivering theorems for each of S_1, \dots, S_n should also suffice for the derivation of such a theorem for S .¹⁹ A theory which satisfies the Mirror Constraint will thus be one whose deductive structure reflects the (causal) interrelations among speakers' sub-competences. When speakers are able to move to understanding of a novel utterance without special explanation, the theory will mirror their ability by supplying the means for deriving an appropriate theorem utilizing only axioms adequate for the specification of the meanings of sentences which they previously understood; and when speakers change their beliefs about the meaning of some sentence, appropriate modifications to the meaning-delivering theorem for that sentence will enjoin revisions in its axiomatic parentage which in turn entail shifts of meaning in exactly those sentences of which they consequentially change their understanding.

Davies himself raises various objections to the Mirror Constraint.²⁰ But he does not raise what I think is the most serious: that it provides no real reason for putting structure into a *semantic* theory. Let it be granted that the interrelations of competence whose reflection the Mirror Constraint seeks to ensure are worth describing. Still, why not describe them directly – why run the dogleg of having them 'reflected' in the deductive articulation of a theory of meaning? There is nothing to prevent a critic of the Mirror Constraint from taking over the syntax and catalogue of semantic primitives incorporated in a theory of meaning which satisfies it. He may then advance a theory of meaning of the infinitary sort canvassed above, adding only a rider to the effect that speakers are generally able to understand novel sentences, provided they involve only familiar semantic primitives, and that changes in their semantic beliefs about a sentence tend to be associated with changes in their semantic beliefs about *all* sentences – or at least all about which they have any such belief – containing some one or more of the semantic primitives figuring in that sentence. Admittedly, such a rider would not be a detailed, or

19 This does not coincide exactly with Davies' formulation, but differs, I believe, in no important respect.

20 For instance, that the constraint provides no guidance to the semantic theorist if the studied language has no actual speakers; and leaves no space for the idea that speakers might fail to know the meanings of sentences which are nevertheless determinate, fixed by syntactic constructions and semantic features familiar to them.

axiomatic, description of the interrelations which the Mirror Constraint would have a theory of meaning reflect. But there is every reason to think that the recursive *syntax* which the theorist adjoins to his infinitary semantic theory would supply the materials for the more specific descriptive task. He needs only to ensure that the syntax itself meets the Mirror Constraint: that when, and only when, speakers' understanding of S_1, \dots, S_n and S are interrelated as described, those ingredients in the axiomatic bases and set of recursions for the syntax which suffice to characterize each of S_1, \dots, S_n as well-formed suffice so to characterize S . There is a different line of thought in Davies' discussion which seems more promising. It depends upon our being willing to entertain the idea that there is an admixture of rational inductive and deductive inference which can take a subject from knowledge of the meanings of a finite set of sentences to knowledge of the meaning of a sentence which is not in that set and is novel to him. It is, Davies writes,

...the possibility of self-conscious, reflective projection of meanings which encourages the attempt to provide a theory of meaning which not only delivers the correct meaning specifications but also reveals how the meanings of sentences depend upon the recurrence of particular syntactic constituents...²¹

Davies proceeds to propose what he calls the *Structural Constraint*: in effect, that if, but only if, it would be possible for someone who knew what S_1, \dots, S_n each mean to proceed, by rational inductive and deductive methods and without further empirical investigation, to knowledge of what S means, the smallest set of axiomatic resources which suffice, in a theory of meaning for the language in question, to furnish meaning-delivering theorems for each of S_1, \dots, S_n should also yield a meaning-delivering theorem for S . The effect of this constraint is that compositionality in a theory of meaning is demanded not by characteristics of actual speakers qua actual speakers but by the nature of an idealized epistemology of understanding. Whatever actual speakers do or do not know, it seems highly plausible that there is such a thing, in certain cases, as rational inference from knowledge of the meanings of the sentences in a particular set to knowledge of the meaning of a sentence outside that set: the cases in question are precisely those where, intuitively, the semantically contributive vocabulary and syntax of the new sentence are all variously on display among the sentences in the set. The effect of the Structural Constraint is that a satisfactory theory of meaning should mirror not the propensities for meaning-projection and revision of actual speakers but those of an ideal speaker, whose every semantic belief is informed by self-conscious rational inductive and deductive inference.

21 Davies, *Meaning, Quantification, Necessity*, p. 56.

In fact it seems reasonable to demand more. A satisfactory theory of meaning should not merely 'reflect' the path that will be taken by the ideal speaker, by discerning structure whenever he discerns structure, but should represent the inferences which he would – or could, qua rational – actually draw. Thus whenever he is able to advance to knowledge of the meaning of S from knowledge of the meanings of S_1, \dots, S_n , the latter knowledge should constitute good evidence for the truth of those axioms in the theory needed to derive the meaning-delivering theorem for S ; and the movement up, as it were, to the axioms and then down to that theorem ought to be the very movement which knowledge of the meanings of S_1, \dots, S_n is deemed to put a rational speaker in position to make. It is not completely clear whether Davies has this stronger interpretation of the Structural Constraint in mind. But it seems to be more satisfactory. The weaker interpretation is apt to make the constraint seem somewhat arbitrary: what question, exactly, would the theory of meaning have to be addressed to in order for it to be necessary that it discerns semantic structure in a group of sentences when and only when the ideal speaker would 'project' amongst them but not necessary that it represent inferential moves which he could, qua ideal, actually make? Under the stronger interpretation, however, the overriding question is clear. It is: granted that it is possible for a speaker to know, or at least to form rational beliefs, about the meanings of utterances whose use he has never witnessed, how in detail might this be possible?

This has to be a good question unless we are utterly sceptical about whether there is ever any *rational* route to an understanding of novel utterances. The project of trying to answer it has the interest that attaches to any programme of reconstructive analytic epistemology. Such projects have been a major current in the history of English-speaking philosophy: whether motivated by sceptical challenges or not, philosophers have repeatedly been drawn to the task of trying to explain how statements of some particular sort – about God, or the material world, or other minds, for instance – could be susceptible to rational cognition (even if our practice is to rely upon criteria of acceptability which fall far short of it). Admittedly, the sceptic about meaning is a somewhat recent entry on the philosophical scene; the project of constructing theories of meaning of the sort we are interested in did not originate as a response to sceptical pressure, and the various forms of scepticism about meaning which have comparatively recently come into prominence in any case concern more basic matters than knowledge of the meaning of novel utterances. Still, the possibility of such knowledge provides the material for a perfectly familiar species of analytical enquiry.

That is the principal of recommendation of this section. There is a recognizably *philosophical* project – at least it ought to be recognizable to

anyone educated in the twentieth-century Anglo-American philosophical tradition – to which constructing a formal theory of meaning would be a contribution. This project has no immediate connections with the quest to explain the capacities of actual speakers of natural languages; the task rather be that of explaining how a complete knowledge of a particular natural language could be a rational achievement. We have no conception of how that might be so unless the rational subject is permitted to discern sub-sentential semantic structure. It follows that such a theory would have to comply with the second ingredient condition of compositionality, as characterized above, that the meaning-delivering theorems for sentences be derived within the theory in a manner which reflects the semantic contribution made by those sentences' constituents. The need for the first ingredient, that the axiomatic basis be finite, is less immediate: it is not evident that a rational being could not be in possession of infinitely many logically independent items of information. But it may be anticipated that if we are concerned with the powers of an ideally rational *human* speaker – so that the finitude of our capacities remains a constraint on the form which the theory should take – the *learnability* of the language – the possibility of a finite but rational creature coming to know the meaning of any particular sentence of it by way of exposure to and projection from the use of finitely many other sentences – will require that only a finitely axiomatized theory will fit the bill.²²

Whatever we conclude, then, about the capacity of a formal theory of meaning to be yoked to the task of explaining actual speakers' abilities, there is an interpretation of the Davidson/Dummett project which promises to allow it to stand independently as legitimate a priori philosophy. This interpretation may or may not accord with part of the intentions of the leading protagonists in the field. But it does to contrive to supply, after all, some sense for Foster's notion that a natural language might have the sort of autonomy which would allow its theoretical description not to be directly a theory about actual speakers' semantic knowledge; and it suggests, in consequence, how the disinflection of the concept of implicit knowledge need not be a precondition for the philosophical health of the project of constructing theories of meaning.²³

22 Davies shows, in fact, that on natural assumptions about what is requisite for the learnability of a language, the Structural Constraint enjoins that a theory satisfying it will indeed be finitely axiomatized. See *Meaning, Quantification, Necessity*, chapter III, section 2.

23 One important reservation about the proposal would concern whether knowledge of the meanings of an appropriate S_1, \dots, S_n could be enjoyed *independently* of knowledge of an understanding of S_1, \dots, S_n , would be possible. The opposing thought, urged on me in discussion by Elizabeth Fricker and suggested by some of the argumentation in her 'Semantic

III

In order for a theory of meaning to be explanatory of the linguistic capacities of actual speakers, there has to be, it seems, a sense in which its axioms are true of them; or, at least, there have to be properties of the speakers for whose description the axioms are needed. It would be an error to suppose that the notion that speakers *implicitly know* the content of such a set of axioms is the only way of meeting this condition. It is evidently a tempting and natural suggestion – witness the widespread use of such ideas in the writings of theoretical linguists. But it is exceedingly difficult to be clear whether it is ultimately coherent.

The most immediate objection is based on the thought that the axioms of a theory of meaning ought to correspond to semantic *rules*. To amplify: such an axiom is supposed to describe the semantically relevant features of an expression; how could it do that if it failed to embody a condition on the correct use of that expression? For the meaning of an expression is essentially something normative; it is, crudely, only because expressions have meaning that there is such a thing as correct, or incorrect, use of them. This normativity does not *per se* conflict with the capacity of the theory to contribute towards the explanation of speakers' linguistic 'creativity' – their capacity to understand novel utterances; the proposal will be that the feature of a speaker which such an axiom can reflect is

Structure and Speakers' Meaning' (see in particular section IV), would be, in effect, that understanding any such S is a *criterion* for knowing the meanings of S_1, \dots, S_n . Whereas the project of describing a basis for a rational inference to the meaning of S has content only if it is possible to be apprised of the basis and yet fail so to exploit it.

Clearly there is no objection of this kind where we are concerned with a *second* language: of course it is possible for an English speaker to know *what* each of finitely many French sentences means without having any *idea why*. The question is whether that distinction, between knowing what a sentence means and why it means it, has content for a first language – can there be such a thing as understanding a sentence of one's only language without understanding its semantic structure? It seems we would have to deny that there can if an explanation is to be possible of why – supposing it is so – sentence understanding has the sort of holistic character claimed by the objection. But notice, in that case, that we should then have a quite different form of response to the problem of motivating compositionality.

A homophonic theory consisting of a recursive syntax plus an infinitary semantic axiom schema, or – for a finite language – a 'firstform' theory consisting simply of specifications of the meanings of all its sentences, would no longer have a claim to characterize what someone who understood the object-language thereby knew. At least, it would not do so with sufficient explicitness: a fully explicit account would have to contain the resources for describing, for each object-language sentence, not merely what it meant but why it meant it.

A fully satisfactory treatment of our topic would have to investigate this through to a conclusion. But I have felt justified in not attempting this here, because of the prospect that the objection, if sustained, would still conserve an a priori motive for compositionality.

precisely his knowledge of it. If we are given, for instance, a statement of the powers of the pieces in chess, there is no tension about supposing that it may serve both to articulate the norms determining what is and is not a legal move and to contribute towards the explanation of players' capacity to recognize the legality/illegality of moves which they have not previously encountered. The double function is secured by the bridging assumption that the players recognize exactly the rules which the statement describes. The salient point is therefore the need for a bridging assumption which hypothesizes practitioners' recognition of the set of norms which the axioms describe.

So what is the objection? Well, if something like a Davidsonian theory is indeed possible for English, it is a mighty iceberg about whose overall shape we have very little idea. We know that it would be a theory of great complexity which would impose a good deal of regimentation on the surface grammar of our language. The handling of tense, adverbs, predicate modifiers, modality, intentionality and even quantifiers are all controversial topics. The tendency, understandably, among researchers in the field has been to stay close to the Tarskian prototype and to the syntax of predicate calculus. But while interesting work has been done, even the most committed would have to allow that progress towards realization of the grand design has not been spectacular. Accordingly, someone who believes that English, or at least a good deal of it, does indeed admit of complete semantic description by means of a compositional formal theory would at present be hard pressed to justify a high degree of confidence that work has proceeded along the right sort of lines, or has been inspired by the most fruitful paradigms. In other words, apart from knowing that it will be difficult to design, will be complex of articulation and is likely to contain at present unforeseen devices, we have *very little* idea what a formal theory of meaning for a natural language of expressive power comparable to that of English would be like. Yet it is the axioms of such a theory which the implicit knowledge proposal would have us regard as *normative* with respect to our linguistic practice. Are there not manifest philosophical difficulties with the idea that our linguistic behaviour should be regarded as informed by our recognition of principles which we cannot state, which played no explicit part in our linguistic training, which will probably involve concepts of great sophistication and technicality, and which we might not recognize even if presented with a formulation? How can a principle function as a rule if those who engage in the practice which it is supposed to regulate have no consciousness of it?

Dummett writes:

What plays the role, within a theory of meaning of Davidson's kind, of a grasp of the meanings of the words is a knowledge of the axioms governing

those words: in our example [that of 'the earth moves'] these may be stated as "the earth" denotes the earth' and 'it is true to say of something "it moves" if and only if that thing moves'. (This latter formulation of the axiom governing 'moves' is stated without appeal to the technical device of satisfaction by an infinite sequence, and is only an approximate indication of what is wanted: but, if we are intending a serious representation of what is known by anyone able to speak English, we cannot literally credit him with an understanding of that technical device.)²⁴

We can sympathize with Dummett's reservations: it is a plausible enough constraint upon the significant attribution of belief, *a fortiori* of knowledge, that a subject possesses the concepts which figure in the content of the belief attributed to him, and it is utterly implausible that ordinary speakers of English should be credited with an understanding, at any level, of concepts like satisfaction, infinite sequence and the star functor. But Dummett's thought here is vulnerable to a simple dilemma. If this sort of technical apparatus is necessary for the development of a theory of meaning for a typical natural language, then speakers may not, on his own admission, be credited with a full implicit knowledge of that theory but only, perhaps, of a part of it; and the claims of the theory to provide a model of how speakers actually achieve an understanding of novel utterances must therefore be severely qualified. If, on the other hand, the technical apparatus is merely a convenience, an explanation is owing of how it may be dispensed with and the axiomatic and recursive basis of the theory developed purely in terms of concepts whose attribution to speakers is altogether more plausible. The promise of a theory of meaning to illuminate actual speakers' linguistic 'creativity' will then depend entirely on the success of this explanation – an explanation which we have, so far, not the slightest idea how to construct.

It will be apparent that there are two special and separate difficulties here which vindication of the notion of implicit knowledge in general would not necessarily resolve. There is the problem of explaining how a *rule* can be implicitly known, how, that is, it can function as a rule – exert a regulative influence – if practitioners are unaware of it. And there is the problem of how contents, putatively implicitly known, may involve concepts for their possessing which there is no direct evidence in practitioners' performance and which may, indeed, be sufficiently abstruse to be beyond their powers – at least their *apparent* powers – of comprehension. However I do not think that either problem is immediately fatal. It is true, of course, that a rule as a possible object of consultation has to be a possible object of consciousness. But it is perfectly obvious that the axioms of a homophonic theory of meaning are not meant to be rules in this

²⁴ What is a Theory of Meaning?, pp. 109–10.

sense since – because they use the expressions which they mention – only someone who already understood those expressions, and hence had no need to consult such ‘rules’, would be capable of consulting them. It does not follow that the axioms cannot be regarded as statements of rule at all, however. It would indeed be fatal to their claim to contain a theory of *meaning* if they could not. But the simple fact is that the characterization of a convention does not have to proceed in terms which could be used to explain the convention to someone previously ignorant of it. It is a convention of English that ‘red’ in its most basic, literal sense, is correctly predicated only of things which are red. Speakers of English who are credited with an understanding of ‘red’ in its most basic and literal sense are thereby credited, *inter alia*, with the intention to uphold this pattern of predication as a matter of convention. There is no better statement of the convention than the one I have given. And it is, at the same time, perfectly useless as an explanation of what the convention is to someone who doesn’t already know it. The dilemma is therefore a false one: the axioms of a theory of meaning do not have to be explanatory, or ‘objects of consultation’, on pain of failing to concern meaning. It suffices that they describe linguistic conventions. And the question of speakers’ implicit knowledge of them is thus, in effect, the question whether speakers may be regarded as implicitly recognizing the conventions which they state.

The proper analysis of the notion of convention is a subtle business.²⁵ It may be that the idea of a convention which is merely implicitly recognized would turn out to put the notion under great strain. But *prima facie* there seems no reason to expect so.²⁶ Whatever the details of a correct analysis, convention is going to turn out to supervene upon people’s intentions, fundamentally the intention to uphold the regularity which the convention prescribes. The crucial question is therefore whether – whether or not they are, or can be brought to be, aware of it – the axioms and recursions of a theory of meaning might succeed in describing a set of linguistic constraints which competent speakers of English may be regarded as intending to uphold. If the answer is negative, the attempt to interpret the relation between speakers and an appropriate theory of meaning for their language as that of implicit knowledge must fail; but if affirmative, that interpretation has a chance of success. In any event it is to assail a man of straw to insist that the implicit knowledge theorist should explain how a principle can be actively regulative of a practice when none of the

participants can profess to know it.²⁷ The issue rather concerns constraints on the ascription of intention. Does it make sense to ascribe intentions to people which they cannot articulate? Or intentions a correct description of which they cannot recognize when given it?

It is important to recognize that what is at issue here is not the propriety of *extending* the notion of intention so as to accommodate the implicit knowledge theorist’s needs. If ordinary practice is to be the guide, it is clear that we *already* use intention, and cognate notions, in ways that can make his position seem quite natural. The attribution of intention is entailed whenever we credit a subject with agency; and we implicitly credit a subject with agency – the capacity for action, in the proper sense of the word – whenever we deem it appropriate to offer *rationalistic* explanations of its performance, explanations which proceed by the ascription of a system of beliefs and desires. Now the fact is – philosophically suspect, or not, as the practice may be – that we go in for simple rationalistic explanations of the behaviour of more intelligent animals. If a dog sets off from his home at roughly the same time each day, no one would think it outrageous to be offered the explanation that he expects his master to be returning home at about that time and wants to see him. Some philosophers (ironically enough, Davidson)²⁸ would argue that any rationalistic explanation of the behaviour of a languageless creature is misconceived. My point is only the descriptive one that such explanations are *commonplace*: in the above example, the dog is implicitly credited with the intention to intercept his master – an intention of which it can neither give nor recognize an adequate description.

Consider this case. It does not happen, but might, that small children could learn to play chess long before they could learn to understand speech – or at least to attain the level of understanding necessary to give or follow explicit descriptions of the rules of chess, or of points in the theory of the game, etc. They learn the moves, let us suppose, by just the sort of patient mixture of drill, demonstration and (inadvertent) reward by which any ordinary child learns the names for colours, or farmyard animals. And then, remarkably, some of them acquire the ability to play not merely legally but well, responding with subtlety and inventiveness to board configurations which they have never encountered before. It would be overwhelmingly natural to credit such children not merely with knowledge of the rules of chess but with the sort of insight into the potentialities of a situation which any good chess player possesses. It is, however, difficult to understand how such insight should be supposed to

²⁵ The *locus classicus* is David Lewis’s *Convention: a Philosophical Study*.

²⁶ I intend no judgement, by this remark, about whether the very strong epistemic conditions involved in Lewis’s original account *would* permit meaning-theoretic axioms of which subjects were unaware to encode conventions. But the crucial question is in any case that about intention to which the text now moves.

²⁷ One cause for complaint about Baker’s and Hacker’s *Language, Sense and Nonsense*, is their predilection for such opposition. There are other causes. See my review, ‘Understanding Novel Utterances’.

²⁸ See ‘Thought and Talk’, Essay 11 of his *Inquiries into Truth and Interpretation*.

function if it is not essentially a faculty of *inference*: inference which goes to work on premises including, *inter alia*, the rules determining the powers of movement and capture of the various types of piece. It seems to me undeniable that, pre-philosophically as it were, we should be quite content to explain such childrens' performance by ascribing to them the knowledge and intentions constitutive of an understanding of, and the practice of playing by, the rules of chess; we would regard them as able to apprehend the implications of those rules for the potentialities of a particular state of play and thereby able to inform their selection of moves with the aim of winning, or at least avoiding defeat.

The parallel with the demands made by the implicit-knowledge interpretation of the theory of meaning, the speakers being credited with knowledge of the axioms and with the capacity to apprehend their more or less remote consequences, is obvious. Indeed the analogy may seem close enough to call the validity of what I have suggested would be our natural response to the hypothetical children into question. But to stress: my point is only that, whether or not the implicit-knowledge interpretation of the theory of meaning is ultimately coherent, the objection to it – for all we have so far seen, at least – ought not to be that it puts an impossible strain on our *ordinary* understanding of notions like knowledge, belief, intention, inference, etc. On the contrary, it is aspects of our ordinary understanding of those notions which make the implicit knowledge interpretation seem natural. If it is to be rejected, the prospect is thus not of excision of an unwarranted extension of our ordinary understanding but of revision of it.

The example of the dog is suggestive in a different way. Since Brentano, it has been the more or less received wisdom in the philosophy of mind that the truth-conditions of ascriptions of propositional attitude are indifferent to reference failure in the proposition in question, but sensitive to the inter-substitution of co-referential expressions. Whether or not these are genuinely semantic (contrast, pragmatic) phenomena, it is open to question whether either is a feature of the sorts of context in which we ascribe beliefs to, e.g. a dog. As far as indifference to existence is concerned, we go in for the ascription of propositional attitudes to animals only with a view to rationalizing, after a fashion, their modes of interaction with items which figure in their experience; since no non-existent items so figure, there is simply no explanatory role to be played by the ascription to them of attitudes to propositions which we can express only by recourse to empty singular terms.²⁹

²⁹ Any apparent counter-example to this claim is going to be controversial and marginal at best. One possibility: if humans, who share a magic-mushroom, e.g., can co-hallucinate – 'Look at that little green man sitting by the window' – there might be circumstances in which a dog would best be described as e.g. barking at such a 'common object' of hallucination. But I can envisage no other circumstances in which we would have cause to use an empty singular term – rather than a quantifier – in ascribing an intentional state to an animal.

Inter-substitutivity is less straightforward. It is natural to report that the dog expects to see his master in the road, unnatural to report that he expects to see Joe Smith, or the high street newsagent. But the latter descriptions strike us as unacceptable only because we are unwilling to impute to the dog any conception of a personal proper name or the institution of newspaper selling. And the fact is that matters stand no better with the relevant concept *master*. A dog's master is, *inter alia*, ultimately responsible for its welfare and ultimately liable for its good conduct. He has the right to move it around (within, e.g. restrictions imposed by quarantine laws), and even to dispose of it. Does the dog know all that? There is a temptation to reply the dog has, as it were, a *thinner* conception of his master – one shorn of institutional trappings and based entirely on the history of interaction between them. I advise anyone who feels this temptation to attempt to specify this alleged conception, bearing in mind that the terms used must no way exceed the concepts which may plausibly be attributed to the dog. I do not anticipate success. The truth, I suggest, is that we have no definite idea what concept we are attributing to a dog in describing its behaviour in this sort of way; and that the (unsurprising) explanation of this unclarity is that we are not seriously attributing a *concept* at all. 'His master' is a term which we use for the convenience of ourselves or our audience; there is no constraint of fidelity to a mode of conceiving employed by the dog. If the man in question was known to most of those present as the high street newsagent, whose arrival was keenly awaited, and if the relation of the dog to this man were of no importance in context, it would not be unnatural to report that the dog had rushed off in the expectation of meeting the newsagent on the road.³⁰

It is, of course, uncontroversial that it is not *always* a paramount constraint on the satisfactory reportage of propositional attitudes that the content-specifying part involve only concepts attributable to the subject. So much has long been recognized. What I am now suggesting is something stronger: it is our practice, in certain cases, to ascribe propositional attitudes in such a way that not only do we not intend the form of words which we use to reflect the modes of conceiving pursued by the subject but, more, there *is* no form of words which, if it were important to us, would suit that purpose. One response would be that this, if true, displays

³⁰ Indeed, why ascribe to a dog so much as a *sortal concept* of man, a conception of his master as a recurrent *particular*, rather than view it as operating a primitive feature-placing scheme of concepts? Note however that if the main claim of the text – that 'there is no constraint of fidelity to [the subject's] mode of conceiving' in such cases – is correct, it does not follow that co-extensive expressions will be unrestrictedly intersubstitutive in the relevant class of contexts. Whether that is so will depend on what *other* (audience- and reporter-related) constraints are in operation. Any purported counter-example to the main claim will therefore have to be shown not to be the effect of other such constraints.

a serious indeterminacy in the beliefs, etc., attributed and so calls into doubt the propriety of the attribution. But a different response is possible: that it may be perfectly proper to ascribe certain propositional attitudes to a subject even though there is no, as it were, *canonical* specification of the content of those attitudes, no specification of their content which exactly captures their content-for-the-subject. Despite the recent concentration of effort on 'folk psychology' and the intentional, I think we are some distance from the insights to motivate either response. But if the second is tenable, the likely abstruseness of the concepts necessary for the formulation of a full-blown theory of meaning need no longer constitute an objection to the implicit knowledge conception. Crudely: if a dog may have beliefs of which there is no formulation save by the use of concepts some of which should not be attributed to the dog, then perhaps speakers of a natural language may have intentions of which there is, again, no satisfactory account save by recourse to technicalities of which they have no concept. I do not know what it is right to think about this. I would urge only that those who would treat the objection from technicality as decisive against the implicit knowledge conception should recognize what they are doing: it is not so much a matter of siding with common sense as taking (to the best of my knowledge) an unargued stance on fundamental questions in the philosophy of mind concerning the proper limits of explanation by the postulation of intentional states and the proper constraints on the reportage of the content of such states.

There are, however, more subtle objections to the implicit knowledge conception. Let it be accepted that the prodigious children could properly be described as implicitly knowing the rules of chess and as possessing an appropriate range of intentions, and a measure of insight, which that knowledge served to inform. Still, the situation is not perfectly parallel to what is required in the case of a theory of meaning. The difference is that the axioms and recursions of a theory of meaning do not relate to correct linguistic practice as the 'axioms' of a theory of chess – i.e. the rules of the game – relate to correct play. Someone who intentionally and in good faith moves his Queen in the manner of a Knight shows that he doesn't know the rules prescribing the powers of movement of the Queen. There is no comparably simple and direct way of showing that you do not know the axiom governing the use, in English, of 'red' or 'elephant'. This is because only a use of a *sentence* makes, as Wittgenstein put it, a move in the language game. Thus an illicit move cannot violate a single axiom; rather, it has to violate a meaning-delivering theorem, and thereby all the axioms and recursive clauses involved in its derivation. Which among these you should then be deemed to be in ignorance of is a matter to be settled by reference to your use of other sentences in the derivation of whose meaning-delivering theorems those same clauses are involved, in the light of holistic constraints.

So much is obvious enough. Why does it constitute a problem? What it shows is that, even if the chess example is deemed to be entirely persuasive of the propriety in general of the ideas of implicit knowledge of rules, and of implicit inference, it cannot commit us to more, in the case of the theory of meaning, than the propriety of the notion that speakers implicitly know the meaning-delivering *theorems*, and can carry out (implicit) inferences from them. Whereas, of course, what needs to be legitimated is implicit knowledge of the axioms; and implicit inferences to the theorems (which, in any case, speakers are likely to know explicitly). The attraction of attributing implicit knowledge of the rules of chess to the fictional children is based on two things: first, their behaviour has all the trappings of intelligence, insight and purpose which would make it virtually impossible for us to regard it as anything but intentional; second, since they behave exactly as if they knew the rules of chess, the kind of rationalistic explanation which viewing their behaviour as intentional demands can hardly do better than ascribe such knowledge to them. The strength of the analogy is that both points apply to linguistic competence too: it manifestly has the richness which invites rationalistic explanation, and – if a theory of meaning is possible at all – the behaviour which would display knowledge of it would be exactly the behaviour constitutive of linguistic competence. But the weakness of the analogy is that behaving, in all respects short of explicit statement, as if one knew a theory of meaning cannot be distinguished from behaving as if one knew its meaning-delivering theorems; whereas there is no proper subset of the theorems of the 'theory' of chess whose knowledge would constitute the ability to play. The suggestion that speakers implicitly know a full theory of meaning for their language thus makes demands on the notion of implicit knowledge which have no counterpart in the chess example.

The additional demands, of course, are precisely what have to be made if the implicit knowledge conception is to provide an explanation of speakers' capacity to understand novel utterances. But is it not a welcome and foreseen effect of the attribution of implicit knowledge of the rules of chess to the children that we thereby secure the means to explain their recognition of the legality, or otherwise, of moves that they have never considered before? Surely. The difference is that the case for attributing implicit knowledge of the rules of chess to the children does not entirely consist in this phenomenon, but can be stated independently of it. The rules of chess comprise the *smallest* theory – (of this particular subject matter; we shall, of course, need to attribute a lot of other information to them, of different sorts) – which we need in order to give the envisaged sort of rationalistic explanation of their behaviour. Moreover, each item of knowledge which we thereby attribute to them has its own distinctive kind of behavioural display. If we seek a theory of meaning with these

same two features, in contrast, we shall wind up with the sort of infinitary axiom schema which figured in the discussion of Foster above. The explanation of the children's ability to judge novel moves in point of legality may thus be viewed as a welcome by-product of an *independently* motivated attribution of implicit knowledge to them. That is not at all the situation with the implicit knowledge conception of the theory of meaning; linguistic 'creativity' here provides the entire *raison d'être*.

The response will be that it cannot be satisfactory just to credit speakers with knowledge of what is stated by perhaps indefinitely many meaning-delivering theorems, some of which concern sentences which they have never encountered, and leave it at that. The question must arise: what is the basis of this knowledge? My point, however, is not that we should discount this question but that the kind of play made with implicit knowledge in the fictional chess case – which was meant to epitomize the strength of the intuitions that underlie the implicit knowledge conception – provides no precedent for the supposition that this question should have a *psychological* answer. There has to be a perfectly respectable scientific question about the sources of our possession of the knowledge which the meaning-delivering theorems of a satisfactory theory of meaning would describe. But there is no a priori reason why the answer to this question should have to proceed via the postulation of further cognitive states. The sought-for finite basis may be better described in non-psychological terms.

There is an argument in Gareth Evans's discussion³¹ which suggests that extending the notion of speakers' implicit knowledge to encompass the axioms and derivations within a theory of meaning would be a definite error. A rat may acquire the disposition to avoid a kind of foodstuff which is poisonous and has caused it sickness in the past. And we might casually ascribe its unwillingness to eat this material – or one that looked/smelt similar – to the belief that it was poisonous. But we should not, Evans urges, let casual language induce casual thought. Beliefs are essentially things which interact with desires and intentions in the production of behaviour. They are also essentially involved in the production of other beliefs. To ascribe a belief is significant only as part of the ascription of a *system* of beliefs. And what behaviour is expressive of a certain belief depends, in general, upon the other ingredients in this system and in the system of the subject's intentions and desires. Thus my belief that a certain substance is poisonous may manifest itself in a literally indefinite variety of ways. I may, like the rat, avoid the substance. But I may also take steps to ensure my family avoid it, or take steps to ensure they don't! I may take small but daily increasing quantities of the stuff in the belief that I can

thereby inure myself against its effects and that background circumstances are such that it may stand to my advantage to have done so. I may take a large quantity if I wish to commit suicide; and a smaller one if I wish to mangle my way out of some obligation. My belief that the substance is poisonous is thus, as Evans puts it, at the service of indefinitely many potential projects corresponding to indefinitely many transformations in my other beliefs and desires. With the rat, in contrast, concepts like the desire for suicide, or malign intent, can get no grip. The 'desires' which we are prepared to attribute to it are restricted, in the present context, to avoidance of distress; and its 'belief' that the substance is poisonous has consequently no other expression than in shunning it.

Evans' point, well made by this example, is that rationalistic explanations of behaviour are so much idle patter unless we are willing to credit the subject with the sophistication of a manifold system of interacting and evolving beliefs and desires, of a degree of organization sufficient to obstruct straightforward dispositional reductions of any particular belief ascription. There is no such obstruction in the case of the rat. Describing it as believing that the substance is poisonous adds nothing to the claim that it has suffered from it in the past and is now disposed to avoid it. If the rat were, e.g. to shift some of the substance to the habitual feeding place of an aggressor, to prevent her children from taking it, and to introduce some of it into the tea cup of the experimenter, on the other hand, we might begin to feel an incentive for serious rationalistic theorizing.

The force of this train of thought becomes apparent as soon as we ask how a defender of the implicit knowledge conception can distinguish those putatively intentional states, whose content he specifies using the axioms of a theory of meaning which he wishes to attribute to speakers, from the sort of dispositions whose behavioural expression is so inflexibly related to them as to disqualify them from the role of components in serious rationalistic theorizing. One of the chess-playing children will standardly manifest his knowledge of the rule governing the powers of movement and capture of the Queen by conforming to it. But other manifestations are possible: he may attempt to correct an opponent who breaks the rule, refuse to play with someone who makes a habit of doing so, or even deliberately break the rule himself as a somewhat unconventional mode of resignation, or by way of a pretended incompetence in the hope of short-circuiting a game he would rather not play. Likewise, someone who is credited with implicit knowledge of a meaning-delivering theorem may express his knowledge in an indefinite variety of ways, including, in appropriate contexts, lying, assent and silence. But the (implicit) knowledge of a meaning theoretic *axiom* would seem to be harnessed to the single project of forming beliefs about the content of sentences which contain

31 Gareth Evans, 'Semantic Theory and Tacit Knowledge'. See especially section III. Compare Davies, *Meaning, Quantification, Necessity*, pp. 83–6.

the expression, or exemplify the mode of construction, which it concerns. Certainly, the precise beliefs which are formed will vary as a function of the content of the other relevant axioms of which a subject is also being supposed to have implicit knowledge. But what is supposed to be the role of *desire*? What is the (implicit?) desire which explains why the subject puts his semantic axiomatic beliefs to just this use, and what are the different uses to which they might be put if his desires were different?

The question draws a complete blank. The case is, in fact, worse than with the rat. We can begin to tell some sort of story – I did so above – of what sort of enrichment and complication of rodent behaviour might enable us to regard the belief that a substance was poisonous as manifested, via a particular kind of behaviour, along with something other than the desire to avoid discomfort. But what is the desire which, in conjunction with the knowledge represented by the meaning-theoretic axioms, is manifested in the formation of beliefs about the meanings of sentences? And what other manifestation might that knowledge have if this desire was different?³² The truth is that the content of ascribing implicit knowledge of a meaning-theoretic axiom would appear to be no more than the ascription of a disposition to form beliefs about the meanings of sentences featuring the expression, or mode of construction, which it concerns: the disposition, precisely, to form beliefs which are appropriately constrained by the content of the axiom. Although Evans allows his discussion to proceed in terms of what he calls 'tacit knowledge', his own response to this train of thought is to abjure any form of intentionalistic construal of the relation between speakers and the axiomatic content of a theory of meaning. Rather, the axioms should indeed be seen precisely as describing certain dispositions which competent speakers have.³³

32 Matters stand quite differently, of course, once the knowledge becomes *explicit*: lying, assent, silence, sarcastic denial, etc. all provide differing modes of expressing it, *modulo* variable contexts and desires. This, I think, is the correct form of reply to John Campbell's point in 'Knowledge and Understanding' about the relative paucity of projects which knowledge, e.g. of the plot structure of *Bleak House*, might be 'at the service of'. In the relevant sense – that of explaining covarying behaviour as other beliefs and desires are varied – such knowledge is indeed at the service of many projects.

33 It is notable that Evans's argument is explicitly directed only against the supposition that speakers *believe* what the axioms state. Earlier we had cause to take seriously the suggestion that *intention* might be the best candidate, from the point of view of the implicit knowledge conception, for the psychological bond between speakers and the contents of the axioms. The proposal was, roughly, that speakers should be credited with whatever (implicit) intentions would suffice to confer the status of conventions on the axioms. Might this make a difference? For there does not seem to be the same kind of holistic flexibility in what counts as manifesting a particular intention which obtains in the case of belief. If the belief that a substance is poisonous may be manifested in any number of ways, among which avoiding eating it is only one – though a usual – case, the intention to avoid eating it, for instance, is manifested by doing just that.

The reader must form his own judgement about whether the point really is fatal to the prospects of any sort of intentionalistic construal of the relation between speakers and axioms. Let me, though, attempt to ensure that he does so in awareness of some limitations of Evans's own positive proposals. I shall pursue Evans's example of a simple language containing just ten singular terms, a, b, c, \dots , and ten one-place predicates, F, G, H, \dots , together with the single sentence-forming operation of singular term-predicate concatenation. The language thus has 100 possible sentences, and allows of a finite but non-compositional truth-theoretic axiomatization consisting of 100 corresponding instances of the T-schema. Call this axiomatization T_1 ; and contrast it with the compositional axiomatization, T_2 which has 21 axioms: ten assigning denotations to the singular terms; ten stipulating satisfaction-conditions for the predicates; and a compositional axiom to the effect that a sentence coupling a name with a predicate is true if and only if the object denoted by the name satisfies the predicate. Evans's negative proposal is that T_2 should not be seen as describing the contents of any sort of intentional states of speakers of the object language. His positive proposal is that it should be seen as describing dispositions which they have; and, crucially, that even when so interpreted, it may be preferable to T_1 .

The immediate question is: *what* dispositions, exactly, does T_2 describe? Evans's own account proceeds in terms of a notion of 'tacit

The suggestion is difficult to appraise in the absence of a detailed proposal. But there is some cause for pessimism. It is, to begin with, an error to suppose that there is a simple analytic connection between the content of an intention and the behaviour which manifests it. There is such a connection, but it is with whatever behaviour *implements* the intention; whereas the intention may be manifested by unsuccessful efforts to implement it, and indeed by any behaviour which the subject believes may (help to) carry it through. Intention, properly so regarded, will accordingly sustain a similar variety of possible modes of expression to that which characterizes belief. Evans's challenge ought therefore still to be good: how is the attribution of implicit intentions to be distinguished from, and justified in preference to, the attribution of dispositions to speak, and interpret the speech of others, in accordance with the meaning-theoretic axioms? Intention is distinguished from a mere disposition by the possibility of misguided attempts at fulfilment and by the subject's adaptability: his capacity to envisage a variety of ways in which it might be fulfilled and to modify his path accordingly. How can these ideas be made to grip in the present case?

In any case, intention cannot be the *whole* story. To be party to a convention is to have both intentions of a certain sort *and* beliefs – beliefs about just what regularities upholding the convention will require to be sustained. In David Lewis's study, for instance, it is necessary, if a regularity is to be conventional, that each of the participants expects the others to sustain it and that everyone prefers to sustain it if the others do (since a solution to a 'co-ordination problem' is thereby achieved.) So the challenge is immediate: how is the putative belief that everyone else will conform to the axioms of a theory of meaning to be distinguished from the *disposition* to form beliefs, as one successfully encounters novel utterances, that their behaviour will, *ceteris paribus*, conform to the requirements of the meaning-delivering theorems for those utterances?

knowledge' – (by way of deference, no doubt, to the free-wheeling use made of intentional terminology by so many psychologists and psycholinguists) – which, in contrast with what the considerations above might prompt us to regard as *genuinely* intentional states, does admit of an apparently straightforward dispositional account. His suggestion is that a speaker U tacitly knows that, for instance, the denotation of *a* is John if and only if he has a disposition such that:

(II ϕ) (II ψ) [if U tacitly knows that an object satisfies ϕ if and only if it is ψ ; and U hears an utterance having the form ϕa ; then U will judge that: the utterance is true if and only if John is ψ].

Likewise a speaker U tacitly knows that, for instance, an object satisfies F if and only if it is bald, if and only if he has a disposition such that:

(II x) (II α) [if U tacitly knows that the denotation of α is x , and U hears an utterance having the form $F\alpha$, then U will judge that: the utterance is true if and only if x is bald.³⁴

These proposals seem more or less inevitable. 'Tacit knowledge' ought to be a disposition which constitutes understanding; and what is it to understand a sub-sentential expression of Evans's simple language except to be disposed to make the right judgements about the truth-conditions of sentences containing it provided one understands the accompanying name or predicate? But there are a number of difficulties.

The first is, once again, that it is not clear how this interpretation of the relation between speakers and the axioms can provide a reason for preferring T_2 to T_1 . The dispositions which T_2 assigns to speakers are dispositions of judgement concerning whole sentences; so why not simply describe them directly by using T_1 ? Evans's answer is that he intends the notion of disposition to which he is appealing to be understood in a 'full-blooded sense': the ascription of a disposition is to be interpreted as the ascription of an underlying state from which the relevant patterns of behaviour, described in the conditional which articulates what the disposition is a disposition to do, (causally) flow. Thus the difference between T_1 and T_2 is that the former ascribes 100 distinct such states to competent speakers of the object language whereas

tacit knowledge of T_2 requires that there should be 20 such states of the subject – one corresponding to each expression of the language which the theory treats separately – such that the causal explanation

of why the subject reacts in the way that he does to any sentence of the language involves two of these states, and any one of these states is involved in the explanation of the way he reacts to 10 sentences containing a common element.³⁵

In Evans's view the claims of T_1 and T_2 to describe speakers' competence may thus, under favourable circumstances, be empirically adjudicated. A satisfactory neurophysiological account of competence would be decisive,³⁶ but even in advance of attaining that, strong evidence for the superiority of T_2 would be afforded by the empirical findings (a) that speakers acquire the capacity to understand so far unencountered specimens from among the 100 possible sentences on the basis of exposure to utterances which contain the relevant constituents; and (b) that when speakers lose competence with any of the sentences – owing to forgetfulness, or disease, or damage – they tend simultaneously to lose competence with all the sentences which feature one, or both of its constituents.

Now, although I think Evans's deference to neurophysiology is mistaken – since it is evidence of types (a) and (b) which would determine our conception of what kind of neurophysiological theory to settle for – and although it is not clear exactly what account of identity and distinctness among (neurophysiological) states should provide the backcloth to his suggestions, the kind of data which he envisages would obviously be highly significant. But the question, of course, is why such data would properly motivate the adoption of T_2 , rather than T_1 supplemented with some appropriate hypotheses, of a non-semantic sort, about the presumed causal substructure of the dispositions which T_1 describes. This is essentially the objection which featured in the discussion of the 'Mirror Constraint' earlier, and Evans's discussion contains, so far as I can see, no answer to it. The requirement that a theory of meaning should both describe the dispositions which the competent display in their handling of whole sentences and reflect the underlying causal structure of those dispositions – as witnessed by the details of their acquisition and loss and, perhaps, by their neurophysiology – provides absolutely no basis for preferring a theory of meaning to a description, or list, of the meaning-delivering theorems, supplemented by claims like

Some single neurophysiological state is involved in the causal explanation of a speaker's competence with any sentence which features the expression *a*.

Why adopt T_2 , or any theory whose axioms have a *semantic* subject matter, if the task is to reflect the *causal* structure of the dispositions which correspond to the meaning-delivering theorems?

34 ' π ' is here a universal substitutional quantifier; and the variables ϕ, α, ψ , and x , have, respectively, the substitution classes of names of predicate expressions of the object-language, names of names of the object-language, predicate expressions of the metalanguage (English) and proper names of the metalanguage (English). Cf. Evans, 'Semantic Theory and Tacit Knowledge', pp. 124–5.

35 'Semantic Theory and Tacit Knowledge', p. 125.

36 *Ibid.*, p. 127.

There is a connection between this point and a peculiarity in Evans's exposition which the alert reader will already have noticed. Why does Evans speak of tacit knowledge of T_2 as involving 20 states of the subject when the axioms of T_2 are 21? The answer is obvious enough. The account which Evans offers of the dispositions which constitute tacit knowledge of the denotations of singular terms and the satisfaction conditions of predicates have the effect that a speaker who possesses them is thereby disposed to attach the proper significance to name-predicate coupling – since he is thereby disposed to attach the proper significance to sentences formed by coupling names and predicates. But this leaves Evans's proposal open to a simple objection. T_2 would be crippled without the compositional axiom, but if the brief of its axioms were *merely* the description of the dispositions which, on Evans's account, constitute tacit knowledge of them, the compositional axiom ought to be redundant. However there is in view no plausible modification of Evans's proposals concerning the dispositions relevant to singular terms and predicates which would need to be supplemented by a separate dispositional account concerning the compositional axiom. So the conclusion has to be that Evans's proposals misdescribe the content of the axioms of T_2 . The conclusion of the preceding argument is therefore reinforced. We can grant that Evans has provided reason why a theory which concerned itself with a description of the dispositions which constituted a speaker's competence might wish to construe some of these dispositions as concerned with sub-sentential expressions. But since any compositional theory of meaning for a typical natural language will incorporate something like T_2 , and since T_2 will not sustain that interpretation of its brief, Evans has provided no reason why we should seek a compositional theory of meaning.

Evans's proposal is apt to seem dissatisfying in a further respect. His account of what tacit knowledge of the denotation of a singular term disposes a subject to do appeals to a prior understanding of what it is to have tacit knowledge of the satisfaction conditions of a predicate; and vice versa. The two sets of dispositions are thus, as Evans acknowledges, 'interdefined'. Why is that not a recipe for vicious circularity? No one can follow Evans's characterization of what it is for U tacitly to know that the denotation of a is John unless he already understands what it is for U to have tacit knowledge of the satisfaction conditions of predicates in the language in question. If he doesn't understand that, Evans's account will plainly be of no avail to him, since it demands a prior understanding of what it is for U tacitly to know – of some arbitrarily selected singular term, which might be a – that its denotation is so-and-so.

This circularity may seem harmless for two reasons. First, it reflects an undoubted feature of our intuitive conception of what it is to understand sub-sentential expressions: to understand a name *is* to have the capacity

to understand utterances in which it figures, provided one understands the remaining constituents and the mode of construction; and to understand the remaining constituents and the mode of construction *is* to have the capacity to understand utterances in which they feature provided one understands the rest of the sentence, which, in the basic case, takes us back to proper names. Second, circularity of this sort need in any case be no objection if the task is not to provide an *introductory* explanation of the concepts in question but to offer some measure of characterization of them.

Both of these points are fair. But the worry is not that the 'interdefinability' of Evans's axiomatic dispositions reflects no feature of our intuitive conception of what it is to understand the constituents of a sentence, but that, naively perhaps, one wants something better in the characterization of a *disposition*. To characterize a disposition ought to be to characterize both what it is a disposition to do and the circumstances under which it will be manifest. Often we settle for very imperfectly precise characterizations of both. But the complaint here is not of imprecision. If, for instance, I characterize the ductility of a metal by reference to certain observable phenomena which occur under background circumstances *including* the possession by the substance of certain further dispositions; and if it then turns out that a characterization of the distinctive manifestations of some of these further dispositions is possible only by reference to background circumstances in which the substances are assumed to be ductile – if that is the best that can be done, the reproach does not seem foolish that I have so far simply *failed to say* what ductility is. Evans's proposals would seem to leave the dispositions which they aim to characterize in this uncomfortable-seeming position. However, I offer the point more as something which someone who wished to advance Evans's account should say something about than as an objection. Perhaps a more sophisticated account of the notion of a disposition would remove the worry; my own suggestion would be that Evans's proposal should have proceeded by reference to states of a different sort – his real interest, after all, is in the underlying 'categorical' bases. But I anticipate.

One final point about Evans's treatment is worth emphasis. He writes ...it is implicit in what has gone before that the notion of tacit knowledge of a [compositional] theory of meaning, explained as I have explained it, cannot be used to explain the capacity to understand new sentences.³⁷

This is because the dispositions which, on Evans's account, constitute tacit knowledge of the axioms of T_2 , e.g., precisely *are* the dispositions to judge correctly the truth-conditions of novel sentences in the language in

37 Evans, 'Semantic Theory and Tacit Knowledge', p. 134.

question. Evans's claim on behalf of a compositional theory of meaning is that it is likely to give the empirically best attested description of what these dispositions are. I think he is right that, even there is no force whatever in the foregoing objections, this is the most that, on his account of the matter, could be claimed. Accordingly, an *explanation* of speakers' 'creativity' would have to consist, for Evans, in an account of how it is that speakers are prone to acquire just these dispositions on the basis of the incomplete and imperfect sampling in which a typical training in the use of a natural language consists.

This still leaves a theorist of meaning with a contribution to make to the explanatory project. Before an account can be given of the aetiology of the relevant dispositions, we need to know what they are. The ability of a learner to understand a novel utterance can, presumably, be made to seem non-miraculous only if the sample of uses which induced in him the dispositions which he thereby exercises themselves involved exercise of corresponding dispositions on the part of those whose speech he witnessed.

Consequently, when a capacity to understand novel sentences is observed, the theorist of meaning has an indispensable role to play in its explanation, since he must exhibit the regularity between the old and the new.³⁸

What is striking about this suggestion is the width of the gulf which it opens between what, on Evans's account, the theorist of meaning should be about and what in practice those philosophers who have taken an interest – none more than Evans – in the project of a theory of meaning have been content to do. One clear implication of Evans's account, for instance, is that the construction of a useful theory of meaning does demand elevation from the armchair. Data are needed about trainees' learning patterns – about just what 'projections' they tend to be able to make on the basis of exposure to just what sorts of sample – and about patterns of loss, before we can so much as form a best guess at the syntactic categories in terms of which Evans's basic dispositions should be described. This is not what has happened. The relevant syntactic categories have been persistently supposed to be, more or less, those which Frege invented; 'regimentation' of the surface grammar of natural language is acknowledged to be inevitable in the construction of a theory of the sought-for kind. I submit that if Evans's account of the project is the right one, this a priori indifference to the *overt form* of many of the utterances which the novice speaker is able 'creatively' to understand is rather strange methodology. Not that the surface/depth grammar distinction may not be amenable to excellent empirical motivation. My point is that philosophical

38 Ibid., pp. 135–6.

theorists of meaning seem to have assumed its propriety without reliance on the kind of data which, if Evans's account of their project were correct, it ought to depend on.

I do not mean to suggest that those philosophers who have set about the Davidsonian project with respect to (fragments of) English have relied on no data which could properly be viewed as empirical. They have relied, of course, on a rich set of intuitions about particular meanings, and the significance of particular constructions, which competent speakers of English tend to share. The point is rather that they have, by and large, relied on no data concerning language acquisition and loss. Admittedly, this may be taken as showing not that Evans's account of how we should conceive the relation between actual speakers and the target theory is altogether misconceived but only that the right account has not greatly impinged upon the consciousness of workers in the field. So it is worth noting, to conclude this section, that there is a proposal, similar in spirit to Evans's but different in detail, which harmonizes rather better with the relatively a prioristic approach that theorists have followed.

As is familiar, certain species of bird display what appears to be a remarkable ability to find their way home from distant and unfamiliar locations. The ability appears remarkable because unless we were allowed to rely on special equipment and knowledge – compasses, charts, the disposition of the stars and so on – we could not emulate it. How do they do it? There are, of course, a number of differences between this problem and that of linguistic creativity. For one thing, there is no analogue of compositionality; no platitude answer, like 'By understanding the words and the way in which they are put together', is in the offing to constrain a satisfactory answer. For another, part of what has to be resolved is the range of sensory cues to which the birds should be thought of as responding – whereas it is taken to be a *datum* that speakers respond to the overt visible or audible structure of a sentence. But what is importantly parallel is that we do not know how to approach the question about, in particular, pigeons unless we are allowed to construct a theory which, like a theory of meaning, serves to articulate possible modes of information processing. We would seek, that is to say, a theory which, if conjoined with supplementary information about features of its novel location which would, according to our best account of a pigeon's sensory apparatus, be discernible by the bird, would serve to issue in theorems whose content would be an instruction about what (sensed object) to fly towards. Of course, the suggestion that actual birds might *know* the content of such a theory would be vulnerable to the principal objection raised above. But it is in any case a suggestion to which we are not tempted; we do not, in setting about devising such a theory, regard ourselves as committed to viewing pigeons as intentional agents. On the

contrary: the idea is to make them intelligible as a sophisticated sort of *mechanism*.

In a way, it is incidental that there are any such creatures. Even if there were not, the question could be posed whether a device could be designed which would 'home' in the way that pigeons actually do. A positive answer to the question would require a demonstration how a mechanism sensitive to certain features of its environment could process the data thereby accumulated so as to be disposed to relocate itself in the appropriate way. At the first stage, this is *entirely* an information-processing problem: it calls, in effect, for the devising of an appropriate computer program. At the second state, the problem would be that of explaining how this program, plus the relevant capacities of sensitivity and movement, might be incorporated into a physically possible device. The sort of understanding of the actual capacities of pigeons which is called for would be achieved exactly when enough was known about them to enable us to understand how in detail they embody such a device. And, of course, there can be no such understanding before we have formed the appropriate theoretical conception of the powers which the device must have. Doing that requires writing the computer program.

Three points are notable. First, devising such a program is not an *empirical* problem. What is sought is an axiomatic theory which, fed with (successive) appropriately formulated descriptions of environments distinct from 'home', will generate (successive) theorems encoding a successful homing strategy. This is a kind of problem which, when sufficiently precisely formulated, can be cracked in the armchair. The corresponding armchair problem for the theorist of meaning is to devise a theory which will take us from a description of relevant features of an arbitrary utterance to a theorem which characterizes its meaning. Second, the theorist will not best serve the next stage of the explanatory project – that of making good the claim that actual human beings embody, as it were, the relevant program – if he produces a theory with an infinitary axiom base. We do not understand what it would be to build a computer which incorporated infinitely many logically independent items of information in its program but no finite axiomatization of them. Simply to postulate that biological evolution can do what we cannot would be to reformulate rather than solve the original problem. And it is in any case unclear what could constitute neurophysiological reason for thinking that a pigeon, or human being, was the living embodiment of such a theory. A finiteness constraint at least thus appears to flow naturally from consideration of the overall character of the explanation which we are seeking. Finally, the connection between the axioms of the theory and speakers' (or pigeons') dispositions is less direct than on Evans's account. A completed explanation along the lines envisaged will of course involve the identification of (presumably

neurophysiological) states which embody the various items of information corresponding to the axioms of the program. But these states need not be individuated, so far as I can see, as (categorical bases for) distinct dispositions; nor, in general, does there appear to be any a priori reason why the correspondence between the axioms and their neurophysiological realizations should be one-to-one.

I claim for this approach only that it may indicate the shape of a better account of the relevance of a theory of meaning to explaining the capacities of actual speakers than can be provided by play with the notion of 'implicit intentional states, or by Evans's dispositionalist account. No doubt it will encounter problems of its own. It is obvious, above all, that clarification is needed of what it is for a system to 'embody' information – clarification which only a philosopher who is unusually well-informed in computational, psychological and neurophysiological science is likely to be able to achieve – and that there has to be, at least initially, a legitimate doubt in any case about the extension of this sort of notion to natural systems. I have wanted to indicate only that the horizon is not empty of all prospect of satisfactorily yoking together the philosophical project of a theory of meaning and the explanation of actual speakers' linguistic 'creativity'.

IV

It remains to draw some conclusions about how our discussion bears on the realism/anti-realism debate, when it is conceived in Dummett's way. The answer, it should now be clear, is: not at all. The anti-realist claim is that nobody may reasonably be credited with knowledge of the truth-conditions of any of a very substantial class of statements. (Precisely what class will depend upon the degree of anti-realism espoused.) The conclusion is then drawn that truth may not play the central role in a comprehensive theory of (statement) meaning – at least not when understood *à la mode réaliste*. The justification for this conclusion is that the theory is supposed to represent the knowledge in which understanding of the sentences of a language consists, which it must be failing to do if it cannot do better than articulate that knowledge in terms of concepts which they cannot have. Now if the discussion of implicit knowledge above had yielded the result that a theory of meaning simply cannot be concerned with the description of speakers' knowledge at all, then the anti-realist critique of (realist) truth-conditional semantics could not take exactly this form. But two points need emphasis. First, what emerged as problematic was the idea of speakers' implicit knowledge of the content of the *axioms* of a theory of meaning – no reason emerged to doubt the propriety of crediting them with implicit knowledge of the content of the meaning-delivering

theorems; and I anticipate that no such reason could be produced which did not demand rejection of the idea of implicit knowledge altogether. Second, if the more radical anti-realist claims about the dubiety of a conception of verification-transcendent truth are correct, we – the theorists – have no business involving that ‘notion’ in any sort of theory, whether conceived as descriptive of the content of object-language speakers’ understanding or not. There has been some curious muddle about this simple point in recent realist commentary.³⁹ So perhaps it is worth emphasizing the obvious: whether or not the theory of meaning is conceived – as Dummett always urges it must be – as a theory of speakers’ understanding, the project is, trivially, constrained by the demand that the concepts which it uses be in good order. Criticism of that particular ingredient in Dummett’s philosophy of language, or highlighting of the non-sequitur involved in the transition from the claims 1) that the meaning of a sentence is what someone who understands it knows, and 2) that the meaning of a sentence is determined by its truth-conditions, to 3) one who understands a sentence knows its truth-conditions,⁴⁰ is therefore entirely futile if what one is trying to do is to protect realist semantics from anti-realist attack. Indeed the anti-realist case has no less bearing on the desirable form for a theory of meaning to take if the project is conceived as having nothing directly to do with the concepts of actual speakers but is concerned entirely with the idealized epistemology of language acquisition, after the fashion of the proposal extracted earlier from Davies’ discussion.⁴¹

II

The Positive Programme

³⁹ An example is Michael Devitt’s chapter 12, ‘Dummett’s Anti-Realism’, of his *Realism and Truth*.

⁴⁰ Cf. *ibid.*, pp. 207–8.

⁴¹ I would like to thank Martin Davies and Elizabeth Fricker for detailed and very helpful comments on an earlier draft of this essay.