

## Article

# *How Can the Theory of Meaning be a Philosophical Project?*

---

CRISPIN WRIGHT

This paper<sup>1</sup> is concerned with the interpretation of a theory of meaning of the kind which, prescinding from their points of disagreement, is associated with the writings of Davidson and Dummett.<sup>2</sup> Such a theory, for a specific natural language, will consist of a syntactic and a semantic part. The syntactic part will divide the atomic expressions of the language into finitely many basic syntactic kinds, and determine on the basis of that division which combinations of such expressions are grammatically well-formed. The semantic part will contain finitely many axioms characterising the meanings of the atomic expressions in such a way as to render the meaning of any well-formed sentence of the language effectively decidable on the basis of its syntactic composition.

As is familiar, Davidson proposed that the sort of recursive characterisation of truth which Tarski showed how to construct for certain formal languages ('Truth and Meaning', Davidson 1984) might serve as the core of such a theory of meaning for a natural language. The Tarskian prototype has, indeed, dominated philosophers' work on the project. But it is of no consequence for the questions with which I am here concerned whether or not that is fortunate. The project is one which, it was and is widely supposed, promises to shed light on a number of philosophically fundamental concepts, and on the notion of meaning in particular (Cf. McDowell 1976 p.72). Still greater claims have been made for it (cf. Dummett 1975, p.97). But it has been generally assumed that, at the very least, its (complete or partial) execution would cast light on such closely interrelated questions as:

How is it possible for speakers of a natural language, all of whose abilities are finite, to understand a potential infinity of sentences?

---

<sup>1</sup> For a more extended treatment, see my 1986.

<sup>2</sup> See especially Essays 1-5 collection in Davidson 1984 and Dummett 1974 and 1975.

How is it possible to learn a (typical) natural language in a finite time?

How is it possible to understand novel sentences never previously encountered (each of these questions is adverted to by Davidson 1984; see Essay 1, p.8, Essay 2, opening paragraph and p.35).

An immediate thought is that these are somewhat empirical sounding questions. Those who have hoped that work on the theory of meaning will bear on them have had it in mind to explain the *actual* capacities of *actual* speakers (usually, of English). How can that be a philosophical project – or, at any rate, a project for the armchair? Is it not the task of empirical psychology to explain any of our actual intellectual capacities, including those associated with linguistic competence?

The question for this paper is under what assumptions that need not be so. It is straightforward enough to see what general form an answer hospitable to the armchair project would have to take. The axioms for a theory of meaning (for English) which is to be explanatory in the desired way will have to describe, or more loosely, reflect, the states of speakers (of English) from which their linguistic competence issues. It can thus be an *a priori* task to devise a theory which, *if* descriptive, or reflective, in an appropriate way, would address the three listed questions; and a task for empirical science to determine whether the *a priori* product is indeed so descriptive, or reflective, of actual speakers' states. Clearly, however, before this cosily collaborative picture can be appropriate, clarification is necessary of just what sorts of states of speakers might correspond to the axioms of an otherwise satisfactory theory of meaning.

Two suggestions on this point have figured prominently in the literature. According to the first, it is appropriate to think of speakers as implicitly *knowing* the contents of the axioms of some such theory and as deploying those contents in the manner articulated by the derivations within the theory of the theorems specifying the meanings of whole sentences (see in particular Dummett 1984, and Campbell 1982, pp.109–12). This is to be understood as a quite serious and literal use of the concept of knowledge. The axioms of the sought-for theory will describe the propositional content of a range of *genuine intentional states* – semantic beliefs – which competent speakers of English have, although unwittingly for the most part<sup>3</sup>, and which participate in the explanation of their behaviour after the fashion of intentional states in general. On this view, the *a priori* task is to formulate a list of axiomatic contents which, *if* believed by speakers would, e.g., empower them to arrive at correct opinions about the meanings of novel English sentences; and the psychological task is to decide which such contents are actually implicitly believed.

The second suggestion holds that the axioms of a satisfactory theory of meaning for English will describe, not the contents of speakers' intentional

---

<sup>3</sup> Since we have little conscious concept of the detail of an appropriate such theory for English.

states, but certain *dispositions* which they have; broadly, dispositions to form beliefs about the meanings of sentences in ways that are constrained by the axioms' contents (see, in particular, Evans 1981). On this view, the form taken by a satisfactory theory of meaning must be sensitive to the presumed *causal* structure of the physical – presumably, neuro-physiological – states which constitute these dispositions.<sup>4</sup>

Each of these proposals raises many involved philosophical issues. But here our treatment must be brisk. In Part I of what follows, I shall outline an objection to the first, due to Gareth Evans (see Evans 1981, Section III; compare Davies 1981, pp. 83–6) which is, I suspect, decisive. In Part II, I shall criticise the dispositional account, to which Evans himself moves in response to his objection. I shall contend that it suffers from certain internal singularities, makes it difficult to understand how even the initial stages of the project can be properly conducted *a priori*, and needlessly abrogates the very ability of a theory of meaning to speak to the three listed questions. In Part III, I shall gesture at what may be an improved account which, if sustained, will legitimate the 'cosily collaborative' picture. But I should stress here that I believe we are still a long way from earning the right to be confident of the prospects of such an account.

## I

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/smelled similar — to the belief that it was poisonous. But we should not, Evans urges<sup>5</sup>, 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 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 incapacitate myself so as to avoid an obligation. My belief that the substance is poisonous is thus, as Evans puts it, at the service of

---

<sup>4</sup> Or underlie them, depending on one's preferred view of dispositions.

<sup>5</sup> I do not know why Evans himself preferred to couch his own account in intentionalist-sounding vocabulary.

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 intentionalistic 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 organisation sufficient to obstruct straightforward 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, on the other hand, 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 an experimenter, we might begin to feel an incentive for serious rationalistic theorising.

The force of this train of thought in the present context 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, and which he hopes 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 intentionalistic theorising. Someone who is credited with implicit knowledge of a meaning-delivering *theorem* may express his knowledge in an indefinite variety of ways relating to the sentence it concerns, 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 proper content of sentences which contain 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 just did so – about 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? 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 fit 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 thus to abjure any form of intentionalist 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.<sup>6</sup>

It is notable that Evans' argument is explicitly directed only against the supposition that speakers *believe* what the axioms state. Might *intention* be a better candidate, from the point of view of the implicit knowledge account, for the psychological bond between speakers and the contents of axioms? The proposal would be, roughly, that speakers should be credited with whatever (implicit) intentions would suffice to confer the status of conventions on the axioms. Could 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 as 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 suggestion, naturally, is difficult to appraise without a detailed proposal. But there is good 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 just the variety of possible modes of expression which characterises belief. Accordingly, Evans' challenge ought still to be good: how is the attribution of the relevant kind of implicit intention 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

---

<sup>6</sup> Matters stand quite differently, of course, once knowledge of the axioms 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 (Campbell 1982) 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 varying explaining behaviour as other beliefs and desires are varied – such knowledge is indeed at the service of many projects.

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 will require to be sustained if the convention is to be upheld. In David Lewis' study (1967) 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 'coordination 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 successively encounters novel utterances, that others' beliefs and intentions will conform to the requirements of the meaning-delivering theorems for those utterances?

## II

The reader must form his own judgement about whether the foregoing considerations really are fatal to the prospects of any sort of intentionalist 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' own proposals. I shall pursue Evans' example of a simple language containing just ten singular terms, *a, b, c, ...*, and ten one-place predicates *F, G, H, ...*, 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*<sup>7</sup> truth-theoretic axiomatisation consisting of 100 corresponding instances of the T-schema. Call this axiomatisation T1; and contrast it with the compositional axiomatisation T2, which has twenty-one 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' negative proposal is that T2 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 T1.

The immediate question is: *what* dispositions, exactly, does T2 describe? Evans' own account proceeds in terms of a notion of 'tacit knowledge',

---

<sup>7</sup> I say that a theory of meaning is *compositional* if it is not merely finitely axiomatised but depends on axioms which characterise the syntax and semantics of the ingredients of a sentence for the manner in which it supplies that sentence's meaning-specifying theorem.

which, in contrast with what the considerations above may 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:

( $\Pi\Phi$ ) ( $\Pi\Psi$ ) [if U tacitly knows that an object satisfies  $\Phi$  if and only if it is  $\Psi$ ; and U hears an utterance having the form  $\widehat{\Phi a}$ ; then U will judge that: the utterance is true if and only if John is  $\Psi$ ].

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:

( $\Pi x$ ) ( $\Pi\alpha$ ) [if U tacitly knows that the denotation of  $\alpha$  is  $x$ , and U hears an utterance having the form  $\widehat{F\alpha}$  then U will judge that: the utterance is true if and only if  $x$  is bald].<sup>8</sup>

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

The first is that it is not clear how this interpretation of the relation between speakers and the axioms can provide a reason for preferring T2 to T1. The dispositions which T2 assigns to speakers are dispositions of judgement concerning whole sentences; so why not simply describe them directly by using T1? Evans' 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 T1 and T2 is that the former ascribes 100 *distinct* such states to competent speakers of the object language whereas:

tacit knowledge of T2 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 (1984, p.125).

---

<sup>8</sup> ' $\Pi$ ' is here a universal substitutional quantifier; and the variables  $\Phi$ ,  $\Psi$ ,  $\alpha$  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). See Evans 1984, pp. 124-5.

In Evans' view, the claims of T1 and T2 to describe speakers' competence may thus, under favourable circumstances, be empirically adjudicated. A satisfactory neurophysiological account of competence would be decisive (Evans 1984, p.127) but even in advance of attaining that, strong evidence for the superiority of T2 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 its 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' ultimate deference to neurophysiology is mistaken – since it is evidence of types (a) and (b) which would determine our conception of what 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 T2, rather than T1 supplemented with some appropriate hypotheses, of a *non-semantic* sort, about the presumed causal substructure of the dispositions which T1 describes. 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 T2, 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?

There is a connection between this point and a peculiarity in Evans' exposition which the alert reader will already have noticed. Why does Evans speak of tacit knowledge of T2 as involving *twenty* states of the subject when the axioms of T2 are *twenty-one*? The answer is obvious enough. The account which Evans offers of the dispositions which constitute tacit knowledge of the denotations of singular terms and of 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 particular names and predicates. But this leaves Evans' proposal open to a simple objection. T2 would be crippled without the compositional axiom. Yet, if the brief of its axioms were *merely* the description of the dispositions which on Evans'

account, constitute tacit knowledge of them, the compositional axiom ought to be redundant. However, there comes to mind no plausible modification of Evans' 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 does not offer a satisfactory interpretation of the contents of the axioms of T2. The conclusion of the preceding argument is therefore reinforced.

Evans' 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' characterisation 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' 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 might seem harmless for two reasons. First, it reflects an undoubted feature of our intuitive conception of what it is to understand subsentential 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 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 characterisation of them.

Both these points are fair. But the worry is not that the 'interdefinability' of Evans' 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 characterisation of a *disposition*. To characterise a disposition ought to be to characterise both what it is a disposition to do and the circumstances under which it will be manifest. Often we settle for very imperfectly precise characterisations of both. But the complaint here is not of imprecision. If, for instance, I characterise 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 characterisation of the distinctive manifestations of some of these further dispositions is possible only by reference to background circumstances in which 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' proposals leave the dispositions which they aim to characterise in this uncomfortable-seeming position. However, I offer the point more as something which someone who wished to advance Evans' 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, to anticipate, would be that Evans' proposal should have proceeded by reference to states of a different sort – his real interest, after all, is in the underlying 'categorical' bases.

One final point about Evans' proposals 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 (1984, p.134).

This is because the dispositions which, on Evans' account, constitute tacit knowledge of the axioms of T2, e.g., precisely *are* the dispositions to judge correctly the truth conditions of novel sentences in the language in question. Evans' 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 if there were no force whatever in the foregoing objections, this is the most that, on his account of the matter, could be claimed. Accordingly, the explanations called for by our original three questions would have to consist, for Evans, in a further 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.

Evans thinks that 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 samples 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 (Evans 1984, pp.135–6).

What is striking about this suggestion is the width of the gulf which it opens between what, on Evans' account, the theorist of meaning should be about, and what in practice those philosophers who have taken an interest in the project have been content to do. One clear implication of Evans' account 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 class 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' 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' account of the project is the right one, this *a priori* indifference to the *overt form* of many of the previously unencountered utterances which the novice speaker is able 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 theorists of meaning seem to have assumed its propriety without reliance on the kind of data which, if Evans' 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' 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. But the resulting situation hardly seems satisfactory.

### III

It seems to me that there is a proposal, similar in spirit to Evans', but different in detail, which harmonises rather better with the relatively aprioristic 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 those posed by the three questions. For one thing, there is no analogue of compositionality; no platitudinous 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, when conjoined with supplementary information about those features of its novel location which ought, according to our best account of a pigeon's sensory apparatus, to 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 – which in itself seems ludicrous – that actual birds might *know* the content of such a theory would be vulnerable to the principal objection raised above. But it is a suggestion to which we are not tempted; we would not, in setting about devising such a theory, regard ourselves as risking commitment 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 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 re-locate 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 programme. At the second stage, the problem would be that of explaining how this programme, 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 programme.

Three points are notable. First, devising such a programme 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 sentence to a theorem which characterises 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 programme – if he produces a theory with an *infinite* base. We do not

understand what it would be to build a computer which incorporated infinitely many logically independent items of information in its programme but no finite axiomatisation 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 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' 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 programme. 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 realization 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' dispositional account. No doubt it will encounter problems of its own. It is obvious, above all, that clarification is urgently 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 every case about the extension of this sort of notion to natural systems. I have wanted to emphasise only that the horizon is not empty of all prospect of satisfactorily yoking together the philosophical project of a theory of meaning and the empirical scientific explanation of actual speakers' linguistic competence<sup>9</sup>.

*Department of Logic & Metaphysics  
St Andrews University  
Scotland KY16 9AL*

*Ms. Accepted 15.1.86*

### **References**

- Cambell, J. 1982: Knowledge and Understanding, *Philosophical Quarterly*, 32.  
Davidson, D. 1984: *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press.

---

<sup>9</sup> I would like to thank Martin Davies and Elizabeth Fricker for very helpful comments on a text of the longer paper mentioned in n.1.

- Davies, M. 1981: *Meaning, Quantification, Necessity*. London: Routledge and Kegan Paul.
- Dummett, M. 1974: What is a Theory of Meaning? In S. Guttenplan (ed) *Mind and Language*. Oxford: Oxford University Press, 1975; What is a Theory of Meaning (II)? In G. Evans and J. McDowell (eds.) *Truth and Meaning*. Oxford: Oxford University Press.
- Evans, G. 1981: Semantic Theory and Tacit Knowledge. In C. Leich and S. Holtzman (eds.) *Wittgenstein: To Follow a Rule*. London: Routledge and Kegan Paul.
- Lewis, D. 1967: *Convention: A Philosophical Study*. Cambridge, Mass: Harvard University Press.
- McDowell, J. 1976: Truth-Conditions, Bivalence and Verificationism. In G. Evans and J. McDowell (eds.) *Truth and Meaning*. Oxford: Oxford University Press.
- Wright, C. 1986 (forthcoming): Theories of Meaning, and Speakers' Knowledge. In Wright. *Realism, Meaning and Truth*. Oxford: Basil Blackwell. Reprinted in S. Shanker (ed). *Philosophy in the United Kingdom*. Forthcoming. Croom Helm.