USE OF THESES

This copy is supplied for purposes of private study and research only. Passages from the thesis may not be copied or closely paraphrased without the written consent of the author.
REALISM, UNDERSTANDING AND TRUTH

by

DREW M. KHLENTZOS

Thesis submitted for the degree of Doctor of Philosophy at the Australian National University

December 1986
Except where otherwise acknowledged,
this thesis represents my own original work.

D. M. Khlentzos

19/12/86
ACKNOWLEDGEMENTS

I am very grateful to all those who have assisted me in the preparation of this thesis. Paul Thom and Professor Genevieve Lloyd offered me good advice on those chapters of the thesis they were kind enough to read as did Dr. William Grey to whom I owe a special debt of gratitude for his many helpful suggestions over the years about writing a thesis. He also read proofs of parts of the thesis and for this I am most thankful. My thanks also to Stephen Langman for reading proofs of parts of the thesis. Dr. Richard Campbell read and offered good comments on a remote predecessor of this work; his interest and sound advice over the years are much appreciated. In the latter stages, Professor Neil Tennant read and offered very helpful comments on an important section of this work. I am very grateful to Jack Lonergan for the many discussions about matters practical as well as philosophical that we have had over the last couple of years.

I owe a profound debt of gratitude to my supervisor, Dr. Peter Roeper, from whom I have learnt so much over many years. His support and encouragement over that time have been unfailing, his tolerance and willingness to correct my mistakes, seemingly endless. My philosophical debt to him is evident in every page of this work.

My greatest debt is to my wife, Philippa, whose love and support sustained me over the years of preparation of this thesis.
ABSTRACT

Realism, as Michael Dummett understands it, is a thesis about the meanings of sentences of a natural language. The Realist's thesis is that the correct model of meaning for those sentences is a truth-conditional one. Dummett has sought to prove that Realism thus defined faces insuperable objections. Those objections centre around the communicability of the Realist's truth-conditions - how, when these truth-conditions are ones that can obtain without speakers being aware that they do, could speakers possibly evince in their linguistic behaviour an understanding of those conditions?

This thesis is an attempt to meet Dummett's arguments against Realism. In the first chapter, I outline Davidson's form of Realism, ultimately disagreeing with his views on the nature of mind (§§1.1.2, 1.4), but endorsing his holistic approach to meaning, truth and interpretation. Meaning for Davidson is a theoretical notion, underdetermined by linguistic usage. Central to Davidson's position is the belief that truth is primitive, a belief which Frege argued for. I argue that Frege and Davidson are right about this and that this doctrine is inconsistent with a correspondence theory of truth (§1.2).

The second chapter sets out Dummett's Anti-Realism, noting Dummett's requirements for acceptable theories of meaning. With one reservation, I accept Dummett's characterisation of Realism (§2.1). Dummett thinks that our use of language is guided by implicit knowledge of a theory of meaning for our language (§2.2), our grasp of all sentences, in particular the undecidable sentences, consisting in a grasp of their assertibility conditions. Essentially defeasible statements present a problem for Anti-Realism (§2.4). I then present Dummett's Manifestation Argument against Realism, questioning some central assumptions of that argument (§2.4). I conclude the chapter with a discussion of Dummett's objections to holistic theories of meaning - some of these are cogent, but others are based upon misunderstandings, I contend (§2.5).

In chapter three, I critically discuss Dummett's idea that truth is a construct from the more primitive notion of correct assertibility. I argue that Dummett's formulation of the latter notion is unsatisfactory (§3.1.1) and that his most convincing argument for truth's arising from assertibility, which is based on our understanding of time and tense, does not, even if sound, prove what he needs to prove if he is to create problems for Realism (§3.1.2). I then examine Dummett's reasons for holding that a theory of meaning must
contain a subpart which pairs truth-conditions with practical recognitional abilities on the parts of speakers (§3.2). I argue that first-person avowals of others require a truth-conditional model for their meanings (§3.3). With this as a counterexample to a global Anti-Realist semantics, I seek to show that Dummett’s Manifestation Challenge can be answered provided one can rebut Wittgenstein’s Private Language Argument.

Although I do not investigate whether it really does so, I suppose with Dummett that Wittgenstein’s argument entails that meaning must be exhaustively manifest in use. I look at the most cogent form of this argument, which is due to Saul Kripke in the final chapter. I contend that Kripke’s semantic scepticism is self-refuting (§5.2).

The penultimate chapter seeks to examine the intuitionistic foundations of Dummett’s Anti-Realism. I begin with a discussion of the intuitionist’s philosophical position, attending particularly to his views on quantification over infinite totalities (§4.1). Dummett and Dag Prawitz have developed a proof-theoretic approach to the meanings of the logical constants (§§4.2, 4.3). Dummett uses a generalised version of the proof-theoretic notion of a conservative extension to press for revision in our Realist-inspired logical practices (§§4.2, 4.5). I examine and reject the most persuasive form of the argument for revisionism (§4.5.1). I argue that theories of meaning based upon the notions of assertibility or deniability cannot explicate the meanings of the logical constants (§4.4) - to do this, the Anti-Realist must develop an acceptable theory of truth; in this connection, Dummett’s suggestions are inadequate (§4.5). I conclude the chapter with a discussion of Dummett’s ingenious attempt to justify deduction (§4.6). This is Dummett at his very best and here I argue that, somewhat in contrast to the overall tenor of my other conclusions, Dummett is absolutely right.
# TABLE OF CONTENTS

## CHAPTER 1

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>§1.1 Is the Mental Anomalous?</td>
<td></td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>§1.1.1 Psychophysical Laws</td>
<td>2</td>
</tr>
<tr>
<td>§1.1.2 Davidson's Decision Theory</td>
<td>11</td>
</tr>
<tr>
<td>§1.2 The Definability of Truth</td>
<td>23</td>
</tr>
<tr>
<td>§1.2.1 Frege's Argument</td>
<td>23</td>
</tr>
<tr>
<td>§1.2.2 Frege and Tarski</td>
<td>34</td>
</tr>
<tr>
<td>§1.3 Truth Theories and Theories of Meaning</td>
<td>42</td>
</tr>
<tr>
<td>§1.3.1 Truth and Meaning</td>
<td>42</td>
</tr>
<tr>
<td>§1.3.2 Indexicality and Truth Theory</td>
<td>52</td>
</tr>
<tr>
<td>§1.4 Radical Interpretation</td>
<td>63</td>
</tr>
</tbody>
</table>

## CHAPTER 2

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>§2.1 Realism</td>
<td>85</td>
</tr>
<tr>
<td>§2.1.1 Summary</td>
<td>99</td>
</tr>
<tr>
<td>§2.2 Meaning, Use and Knowledge</td>
<td>102</td>
</tr>
<tr>
<td>§2.3 Decidability and Truth</td>
<td>118</td>
</tr>
<tr>
<td>§2.4 Dummett's Argument against Realism</td>
<td>130</td>
</tr>
<tr>
<td>§2.4.1 Summary</td>
<td>140</td>
</tr>
<tr>
<td>§2.5 Holism</td>
<td>143</td>
</tr>
</tbody>
</table>

## CHAPTER 3

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>§3.1 Truth, Assertion and Assertibility</td>
<td>160</td>
</tr>
<tr>
<td>§3.1.1 Dummett on Assertion</td>
<td>160</td>
</tr>
<tr>
<td>§3.1.2 Truth and Assertibility</td>
<td>167</td>
</tr>
<tr>
<td>§3.2 Modest and Full-Blooded Theories of Meaning</td>
<td>177</td>
</tr>
<tr>
<td>§3.2.1 Dummett's requirements on an adequate theory of meaning</td>
<td>177</td>
</tr>
<tr>
<td>§3.2.2 Dummett's objections to modest theories of meaning</td>
<td>182</td>
</tr>
<tr>
<td>§3.3 Other Minds</td>
<td>191</td>
</tr>
<tr>
<td>§3.3.1 The problem for Anti-Realism</td>
<td>191</td>
</tr>
<tr>
<td>§3.3.2 Realism and the conscious experience of others</td>
<td>201</td>
</tr>
<tr>
<td>§3.4 Manifesting Understanding</td>
<td>210</td>
</tr>
<tr>
<td>§3.4.1 Realist replies to Dummett</td>
<td>210</td>
</tr>
</tbody>
</table>
### CHAPTER 4

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>§4.1</td>
<td>Intuitionism</td>
<td>229</td>
</tr>
<tr>
<td>§4.2</td>
<td>Proof Theory and Anti-Realism</td>
<td>246</td>
</tr>
<tr>
<td>§4.3</td>
<td>Model Theory, Proof Theory and the meanings of the logical constants</td>
<td>257</td>
</tr>
<tr>
<td>§4.4</td>
<td>Anti-Realist Theories of meaning</td>
<td>280</td>
</tr>
<tr>
<td>§4.5</td>
<td>Anti-Realist conceptions of Truth</td>
<td>302</td>
</tr>
<tr>
<td>§4.5.1</td>
<td>Revisionism</td>
<td>314</td>
</tr>
<tr>
<td>§4.6</td>
<td>Can deduction be justified?</td>
<td>322</td>
</tr>
<tr>
<td>§4.6.1</td>
<td>The Paradox of inference</td>
<td>322</td>
</tr>
<tr>
<td>§4.6.2</td>
<td>The justification of fundamental logical laws</td>
<td>331</td>
</tr>
</tbody>
</table>

### CHAPTER 5

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>§5.1</td>
<td>Kripke's Sceptic</td>
<td>343</td>
</tr>
<tr>
<td>§5.1.1</td>
<td>The Sceptical Paradox</td>
<td>343</td>
</tr>
<tr>
<td>§5.1.2</td>
<td>Kripke's arguments against the Dispositionalist Thesis</td>
<td>348</td>
</tr>
<tr>
<td>§5.1.3</td>
<td>The Sceptical Solution</td>
<td>350</td>
</tr>
<tr>
<td>§5.2</td>
<td>'Sceptical Solutions' and Conventionalism</td>
<td>355</td>
</tr>
<tr>
<td>§5.3</td>
<td>Dispositions, Intentions and Meaning</td>
<td>374</td>
</tr>
<tr>
<td>§5.3.1</td>
<td>Constraints on an acceptable answer to Kripke</td>
<td>374</td>
</tr>
<tr>
<td>§5.3.2</td>
<td>The Dispositionalist Thesis</td>
<td>378</td>
</tr>
<tr>
<td>§5.3.3</td>
<td>Intentions and Meaning</td>
<td>388</td>
</tr>
</tbody>
</table>
§1.1: IS THE MENTAL ANOMALOUS?

Introduction

My aim in this thesis is to defend Realism against the sustained and powerful attack Michael Dummett has launched against it on behalf of his Anti-Realist. Dummett's criticisms of the doctrine he calls Realism concern the theory of meaning developed by Donald Davidson which is fashioned around a Tarskian theory of truth that takes classical truth as its central explanatory concept. The first step in understanding Dummett's criticisms is, accordingly, to understand their target. This is my aim in this chapter.

It might seem surprising that I should begin with Davidson's views on psychophysical laws in an effort to understand his philosophy of language but I think that one cannot hope to sensibly evaluate Davidson's views on radical interpretation until one has some understanding of what he takes the mental to be like. To do that, one needs to take as close a look at his writings on the mind or on action as one would take at his writings on truth and meaning. This is not surprising when one thinks about it since Davidson continually stresses the holistic nature of the mental and its interanimation with the semantic realm.
I have another motive for starting at this point. I do not believe that one can develop an adequate theory of meaning without at least implicitly committing oneself to substantive theses about the nature of the human mind, for linguistic utterances are intentional actions and to achieve a theoretical understanding of their nature we need to have some background theory, however inexplicit, of intentional action in general. Davidson has a very well-developed theory of action, one which I find extremely attractive. The Realist needs such a theory if he is to meet the challenge put to him by the Anti-Realist. To a large extent, it is because it presents us with a palpably inadequate view of the mental that Anti-Realism as a position in the theory of meaning has to be rejected ... or so I argue in what follows.

§1.1.1 Psychophysical laws

Donald Davidson has claimed that there can be no strict psychophysical or psychological laws. His argument for this striking claim seems to turn on the fact that rationality, as that essentially normative concept which controls our attribution of propositional attitudes, cannot be accommodated by physical theory. Thus he writes:

"In inferring the whole system of beliefs, desires etc from the evidence, we necessarily impose conditions of coherence, rationality and consistency. These conditions have no echo in physical theory which is why we cannot have more than rough correlations between psychological and physical phenomena"¹

¹ 'Psychology as Philosophy', at p.231 in Essays on Actions and Events, Clarendon Press, Oxford, 1980, pp.229-239. All page references to Davidson's writings, except where otherwise stated, will be either to that work or to Essays on Truth and Interpretation, Clarendon Press, Oxford, 1984, henceforth either Actions and Events or Truth and Interpretation.
In what follows, I shall be concerned almost exclusively with the question of whether The Argument from Normativeness as I shall call it really does show such laws to be impossible.2

First, a qualification and a reservation. The qualification concerns the scope of Davidson's anomalist theses. Davidson delimits the mental by the criterion of intentionality. By itself this is uncontroversial since everything then depends upon how we articulate 'intentionality'. But Davidson makes it clear that for some event or state or property to be 'intentional' is for it to be describable in the language of the propositional attitudes ... and this brings me to my reservation: surely this is too narrow an understanding of 'intentional' or 'mental'? For it doesn't even seem to count pains as mental. But then what are we to make of a generalisation such as 'S is in pain iff S's C-fibres are firing'? Is this not a psychophysical generalisation by Davidson's reckoning?

This seems to make Davidson's original anomalist theses rather tame tigers.

For now the claim appears to be that there are no laws connecting propositional attitudes to physical states or conditions or to other propositional attitudes ... and anyone remotely convinced by Quine's writings on these matters will have no objections to that claim.

If he is to make his claim that:
'there are no strict deterministic laws on the basis of which mental events can be predicted or explained'3

2 Davidson himself seems to see this argument as the crucial one – witness his recent claim in his 'Reply to J.J.C. Smart' at p.245 of Hintikka, M and Vermazen, B (eds) Essays on Davidson: Actions and Events, Oxford University Press, Oxford, 1985:
'The basic reason the mental concepts connected with propositional attitudes cannot be incorporated in a system of exceptionless laws is the normative character of these mental concepts.'

3 'Mental Events' p.208 in Actions and Events.
as general as he clearly wishes it to be, Davidson needs a criterion for mentality broader than his official propositional attitude one. In another sense, however, his criterion is too broad, for he imposes no constraints on the types of descriptions deploying intentional terms that can be used to individuate events. Events are particulars in the causal network according to Davidson and he tells us that:

'... an event is mental if and only if it has a mental description ... ' ... (1)

Intuitively, one would have classified an event involving the collision of two stars as a paradigmatically non-mental event. Yet Davidson himself shows how to so describe such an event using propositional attitude predicates that it is classed as a mental event by his criterion at (1) above. Thus if 'P(x)' is a predicate true of one such stellar collision C at t and of ones similar to it, we can individuate C from all others if we but know that it was the very time that Jones noticed a pencil roll across his desk. Thus, since the description: '(x)(P(x) & x stands in the spatio-temporal relation R to Jones' noticing a pencil roll across his desk)' deploys a propositional attitude predicate to identify C, it follows by Davidson's criterion in (1) that C is a mental event.

Mark Johnston contends that Davidson is committed to a parallel consequence regarding physical descriptions of events. As for the mental case, events are only physical if they have a physical description. Now suppose that the Big Bang, to use Johnston's own example, has a certain physical description uniquely true of it - e.g. '(x)(Q(x))'; then the mental

4 loc. cit. p.211.
5 loc. cit.
7 'Mental Events' p.211.
event which is Jones’ noticing a pencil roll across his desk and which, like all
events, can be causally or spatio-temporally related to the Big Bang, can be
described physically as:

\[(sy)(y \text{ stands in the causal/spatio-temporal relation } R^* \text{ to } (x)(Q(x)))\]

which, by Davidson’s reckoning suffices to make it a physical event. Thus:
‘Bizarrely, a minimal condition on the connectedness of the causal and spatio-temporal
combined with Davidson’s criteria for mental and physical events suffices to guarantee not only
that all events are mental but also that all events are physical.’

But then Davidson’s monism would be trivialised.

I think Johnston has saddled Davidson with a thesis to which he is not
obviously committed. Davidson does claim that ‘a clearly acceptable
criterion for the identity of events’ is that:
‘events are identical if and only if they have exactly the same causes and effects.’

Davidson also thinks that this criterion of event identity explains why we:
‘so often identify or describe events in terms of their causes and effects.’

But it does not seem to be any part of Davidson’s claim that spatio-temporal
relations alone can serve to identify an event and this thesis has little
independent plausibility.

Johnston’s quick reductio therefore fails. Yet Davidson’s criterion for the
mentality of events does not seem to really be acceptable - there must be
more to an event’s being mental than its being describable in the language of
the propositional attitudes. The criterion seems at once too loose and too

---

8 ‘Why Having a Mind Matters’ p. 411.
9 ‘The Individuation of Events’ at p. 179 in Actions and Events, pp. 163-180.
10 loc. cit.
11 Davidson’s actual method of event individuation is always to use gerundives ‘A walking’, ‘A
flipping of a switch’ etc. These play the same role in the identification of events as sortals play in
the identification of material objects.
restrictive. The natural presumption is that it is a feature of the event or state itself rather than our description of it which constitutes it as mental.

Jaegwon Kim has reconstructed Davidson's Argument from Normativeness in a way which Davidson would, I feel, accept. Kim writes:

'The skeletal structure of what I take to be Davidson's principal argument can ... be exhibited as follows:

The mental system has a certain essential characteristic \( X \) and the physical system a certain essential characteristic \( Y \), where \( X \) and \( Y \) are mutually incompatible. Laws linking the two systems, if they exist, would "transmit" these characteristics from one system to the other, leading to incoherence. Therefore, there can be no laws connecting the mental with the physical so long as the two systems are to retain their distinctive identities.'

Kim takes it that the 'essential characteristic' of the mental for Davidson is rationality and that the constitutive principle of the physical is 'the absence of rationality'. To be more specific, the ascription of contentful mental states to an agent is governed by the principle that his total system of propositional attitudes must be maximally coherent and rational; no parallel principle about the ascription of physical states to a physical system seems to hold.

Let us call this principle the Rationality Maximization Principle (RMP). The mental is supposed to fall under the jurisdiction of (RMP). Kim then proceeds to demonstrate how the jurisdiction of (RMP) would be threatened by the existence of psychophysical laws:

(i) Suppose that on the available evidence, the attribution of either of the mental states \( m_1 \) or \( m_2 \) is warranted and that (RMP) enjoins a choice of \( m_1 \) over \( m_2 \).

(ii) Suppose there are neural states \( n_1, n_2 \) which are nomologically co-extensive with \( m_1 \) and \( m_2 \) respectively - that is, we have laws affirming that as a matter of fact, \( m_1 \) (\( m_2 \)) occurs in organism 0 iff \( n_1 \) (\( n_2 \)) occurs in 0.

(iii) Because \( n_1 \) and \( n_2 \) are theoretical states of a physical theory (to wit, neurophysiology), they have attribution conditions governed by that theory.

(iv) So, if \( C_1 \) is the attribution condition for \( n_1 \), we will have:

\[ \text{true} \text{ if } C_1 \text{ obtains, } n_1 \text{ occurs} \]

(v) But the psychophysical law in (ii) tells us that

\[ \text{true if } m_1 \text{ occurs iff } n_1 \text{ occurs} \]

---

whence \( (3) \) \( \Box ( \text{if } C_1 \text{ obtains, } m_1 \text{ occurs}) \)

(vi) But \( (3) \) tells us that when a certain set of physical conditions holds a specific mental state is of nomological necessity instantiated in \( O \).

(vii) So \( (RMP) \) stands in danger of being pre-empted since the determination of whether these physical conditions hold is not subject to this principle but rather to those stipulated by neurophysiology.

\( (RMP) \), let us suppose, enjoins the ascription of \( m_1 \) rather than \( m_2 \) to Jones but the neurophysiological evidence suggests that Jones is in state \( n_2 \). Kim's thesis is that the nomological status of \( (3) \) is inconsistent with regarding \( (RMP) \) as the governing principle for our attributions of mental states - the assertion that \( O \) must be in state \( m_1 \) if it is in state \( n_1 \): no comparable modal status attaches to our de facto judgements about Jones' mental states as these are controlled by \( (RMP) \).

There is an ambiguity in this position. Is Kim saying that \( (RMP) \) is simply the constraint which \textbf{we} as ascribers of mental states to Jones must obey or is he saying something more - that the relations between Jones' states \textbf{themselves} must be such that rationality is maximized?

Prima facie, there is no obvious reason why the constraints of rationality upon \textbf{us as ascribers} of attitudes to Jones should determine the rationality or otherwise of the states ascribed. It ought to be uncontroversial that agents who attribute mental states to others and to themselves do so in obedience to the constitutive ideal of rationality and we ought to agree with Davidson, if only for the sake of argument, that insofar as they do this, agents take themselves to be theorizing about the causes of the subject's actions. Jones' actions will be explicable, we think, if when his reasons are set out in the form of a practical syllogism, the course of action he took is seen to follow as a deductive consequence of those reasons.
Deviations from the ideal norm of rationality are correspondingly difficult to comprehend. We can largely thank Davidson for making this intuitive idea so appealing to us. So we can agree with Davidson's remark that: 'If attitudes can be identified at all, then, they must be found to be largely consistent with one another (because of their logical properties), and in tune with the real world (because of their semantic properties).'

But the question of scientific interest is surely whether such attitudes are required to explain the behaviour of an agent. Why should our folk theorizing about the causes of behaviour be any different from the scientific viewpoint from our folk theorizing about the causes of other natural phenomena? That is, whilst we can agree that the injunction to maximize rationality and coherence in our ascriptions of mental states to others and to ourselves is one we cannot avoid if we are to produce causal explanations and predictions of the behaviour of agents in the absence of detailed neurophysiological evidence about those causes, we do not and ought not to regard it as inviolable. Whatever the social, ethical and cultural pressures upon us to see ourselves as rational might be, it is an empirical question, or so the physicalist claims, to what extent we actually are rational.

Davidson clearly disagrees with this last claim. Kim does also and produces an argument purporting to show that to allow psychophysical laws to underpin maxims of rationality is to remove the attitudes from the domain of rationality.

Thus suppose that q is an obvious logical consequence of p and (RMP) enjoins us to attribute q to S whenever we attribute p. Suppose now that some psychophysical law L says that it is nomologically necessary that if S is in the mental state of believing p he will be in the brain state B₁ and that if he is in the mental state of believing q, he will be in the brain state B₂. Then, even
though inferences involving counterfactuals are uncertain, it seems right to infer that if $S$ were in the state $B_1$, he would be in the state $B_2$.

Kim’s worry is that the only plausible way to explain why this relation of physical dependence held would be by some more fundamental physical law. But then, according to Kim, (RMP) will have been pre-empted as the ground for the original relation of counterfactual dependence between the states of believing $p$ and of believing $q$ and the concept of belief will have been removed from the domain of rationality.

To a physicalist this will seem the exact reverse of the truth. By showing that there is a physical explanation in terms of ultimate physical laws as to why $S$ is in state $B_2$ whenever he is in state $B_1$ we have not pre-empted (RMP) at all— to the contrary, we have shown why in this instance its predictions are correct, why in this instance we can afford to trust it. If humans are, as we certainly take them to be, minimally rational, then the more obvious deliverances of (RMP) will surely correspond to something in the physical facts. To use an example of Brian Loar’s, physical state types $x$ and $y$ will be related as the belief that $p \& q$ and the belief that $\neg p$ are counterfactually related just when it is the case that if $x$ were to occur, $y$ would not. Loar sums up the physicalist response to Davidson nicely when he claims: 'It is apriori that if certain states are to be counted as beliefs and desires they must satisfy the constraints of rationality. But that they do satisfy them can be as contingent as you like.'

What reason is there for believing that the mental is the autonomous domain Davidson assures us it is? It might be thought that in the case of sensations such as pain, we should always accept the mental criterion of how it feels to Jones rather than the physical criterion of which of his neurons are firing as

---

the decisive criterion of whether or not Jones is in pain. So there can be no prospect of a mature neurophysiology displacing the mental here.

I think the physicalist should concede this in the case of sensations. It is only to the extent that the firing of C-fibres explains Jones' uncomfortable sensations that we are prepared to admit it as a possible independent criterion for his being in pain. Having established as best we can the reliable behavioural signs of Jones' being in pain, we can then ascertain the corresponding physical states causally responsible for his experience, realising all the while that our ultimate criterion is the experiential one.

Is this an unacceptable intrusion on the closure and comprehensiveness of physical theory? Surely not, for pain was defined as that state causally responsible for certain repellant experiences.

The physicalist will be far more wary of attaching much credence to claims of first person authority with respect to beliefs, desires and the other psychological attitudes since here, unlike the case of pain where it seems undeniable that the awfulness of the sensation is causally efficacious, it is quite unclear whether folk-theoretic posits really do 'carve Nature at the joints'. The physicalist ought, I think, to follow J.J.C. Smart's lead here:

'Reliable prediction of human behaviour can come only from scientific psychology, which avoids reference to intentional states. Psychology must become neurophysiological, or at the least it must refer to functionally defined entities whose precise neurophysiological embodiment in any particular case may be unknown. This scientific psychology will not attempt to explain human actions as flowing from intentional states, but will cut human behaviour up in very different ways from those of common sense.'

Far from providing a decisive demonstration of the impossibility of psychophysical or psychological laws, it seems that The Argument from Normativeness simply has nothing to say to the cognitive scientist or neurophysiologist convinced of the existence of such laws, though sceptical

15 'Davidson's Minimal Materialism' p.178 in Hintikka, M and Vermazen, B (eds), pp. 173-182
that these will be found to hold at the level of propositional attitudes. In the
course of such a dispute, the complaint that a decision not to accept the
criterion of the mental in terms of the vocabulary of the propositional
attitudes amounts to a 'change of subject' just seems question-begging—
particularly so when Davidson’s own criterion of the mental seems, as we've
seen, to be less than adequate.
There must be more to Davidson’s argument than that, he surely couldn’t be
resting his whole case for the anomalism of the mental on the intution that
the mental constitutes an autonomous domain whose constitutive principle
of rationality differs in kind from anything to be found in the physical realm.
I think there is more to his argument, but it is to be found in his writings on
decision theory rather than in his writings on philosophical psychology.

§1.1.2 Davidson’s Decision Theory.
Davidson seeks to elucidate the notion of intention by first characterising
intentional action. That an agent has certain psychological attitudes in the
light of which an action appears reasonable does not suffice to explain why
he performed that action, he contends. For A can have the requisite reasons
R for Øing and yet not Ø or Ø for some other reason. In order for R to explain
A’s Øing it must be the case that A Ød because of R.
By constructing A’s ‘primary reasons’ for acting, then, we’re eo ipso
uncovering the causes of his action according to Davidson. Primary reasons
consist of some attitude disposing the agent in favour of some course of
action (a pro-attitude) or a related belief or both.
However, A’s primary reasons will rationalise his action only when that
action is described in certain ways. In giving A’s reasons we may thus infer
his intention in acting although the converse does not hold. Davidson’s hope
in 'Actions, Reasons and Causes' was to define intentional action non-circularly as: those actions done for a reason. He correspondingly sought to effect an ontological reduction of intention to belief-desire pairs. The question is then 'How do reasons cause actions?'

The simplest model is the Aristotelian practical syllogism in which an action is deductively related to the reasons supporting it as a conclusion of a syllogism to its premises.

To take a simple example:

\(D_m\) Any action of mine that will slake my thirst is desirable.
\(B_n\) Drinking a glass of water will slake my thirst.
\(A_k\) Drink a glass of water.

Here, \(D_m, B_n\) constitute A's primary reasons; \(A_k\) is explained by \(D_m, B_n\) and deduced from it. But there is a problem in generalising this simple model.

For it cannot account for irrational but nonetheless intentional action such as Akrasia nor for actions arising out of moral conflict or conflict in choices generally. The reason is that courses of action can be deduced from the premises that an agent holds true in such situations that conflict with each other. Take the politician's dilemma:

---

16 As Michael Bratman observes in 'Davidson's Theory of Intention' in Hintikka, M and Yermazen, B (eds) p.13:
'This conception of intentional action makes no essential appeal to any distinct state or event of intending... that intervenes between my desire-belief pair and my action. The intentionality of my action lies, rather, in its relation to my desire-belief pair.'

17 Cf. Davidson's comment at 'Actions, Reasons and Causes' p.16 in Actions and Events:
'The practical syllogism exhausts its role in displaying an action as falling under one reason; so it cannot be subtilized into a reconstruction of practical reasoning, which involves the weighing of competing reasons.'
Any action strengthening the defence of Australia is desirable. Any action making Australia a nuclear target is deplorable.

Permitting US bases in Australia strengthens her defence. Permitting US bases in Australia makes her a nuclear target.


Aristotle's model takes no account of our weighing consequences or of our evaluating the desirability of different courses of action, according to Davidson. Nor does it allow that the conclusion of a piece of practical reasoning might be the formation of an intention rather than the performance of an action. A new model is required. Davidson turns to Hempel's model of probabilistic reasoning at this point:

If the barometer falls it almost certainly will rain. If the sky is red at night it almost certainly will not rain.

The barometer is falling. The sky is red tonight.

It almost certainly will rain. It almost certainly will not rain.

Hempel argued that the modal operator, 'almost certainly', modified the 'if...then' connective rather than the consequent of the conditional in which it is embedded in \((P_1)\) and \((P_2)\). Thus \((C_1)\) and \((C_2)\) cannot be detached from their respective pairs of premises since the connective lies within the scope of the modal operator. Hempel reconstructed the reasoning as follows:

\[
\begin{align*}
(P_1) & : \text{pr}(R, F) \\
(E_1) & : \text{F} \\
(C_1) & : \text{pr}[R, \text{pr}(R, F) \& \text{F}] \\
(P_2) & : \text{pr}(-R, S) \\
(E_2) & : \text{S} \\
(C_2) & : \text{pr}[-R, \text{pr}(-R, S) \& \text{S}]
\end{align*}
\]

Here \(\text{pr}(R, F)\) is a representation of "that the barometer falls probabilises that it will rain". In predicting the weather we take into account \((E_1)\), \((E_2)\) and all our available evidence. If we let \(E\) be all relevant evidence, then neither \(\text{pr}[R, E]\) nor \(\text{pr}[R, -E]\) can be inferred from the above arguments.
Davidson takes the evaluation of actions as desirable or not as modifying the quantifier in the major premise of the practical syllogism in just the way that the probability operator modified the 'if...then' connective in the major premise of the probabilistic syllogism. The result is that desirability or undesirability is a characteristic of kinds of actions - it is desirable that: any action of mine be a slaking of my thirst, rather than: every action of mine that is a slaking of my thirst is desirable. The latter is presumably false for most agents when the action is one of drowning. So my pro-attitude is not that I slake my thirst simpliciter, but that any act of mine would be desirable inssofar as it is a slaking of my thirst.

Instead of treating judgements of desirability as operating upon single sentences, Davidson defines a functor which maps pairs of sentences related as moral judgement and ground into a sentence expressing the judgement of desirability. By taking two arbitrary actions, one of which evinces the feature F that the other lacks and declaring that the first is 'prima facie better than' the second because of this, Davidson achieves the same effect as saying that any F-type action is desirable.

One horn of the troubled politician's dilemma can then be represented thus:

\( (D_1^*) \): pf( x is better than y, x is a defence of Australia and y is not)
\( (B_1^*) \): a is a defence of Australia and b is not
\( (C_1^*) \): pf( bab, (D_1^*) and (B_1^*)).

The other horn:

\( (D_2^*) \): pf( byx, \neg Ty \& Tx)
\( (B_2^*) \): \neg Tb \& Ta
\( (C_2^*) \): pf( bba, (D_2^*) and (B_2^*)).

The 'pf' operator thus relates sentences - one expressing a moral judgement, the other the grounds for that judgement.
A judgement that is especially important in serving as an action-guiding principle is $pf(Bab,E)$ where $E$ is all the evidence relevant to the issue of whether to permit or prohibit US bases. In making his decision, the politician will move from a $pf$-judgement to an unconditional judgement: Bab. He will then do a if he does either a or b intentionally, since Bab corresponds to his intention to perform a rather than b.

So Davidson’s revised schema for practical reasoning looks like this:

\[
D_1, D_2, \ldots, D_m, \text{ Agent’s pro-attitudes} \quad \text{these are formulated} \\
B_1, B_2, \ldots, B_n, \quad \text{beliefs} \quad \text{as $pf$-relations in the premises of the} \\
\quad \text{syllogism.} \\
\]

\[
C \quad \text{A prima-facie judgement.} \\
I \quad \text{An unconditional judgement.} \\
A \quad \text{The action done with the intention } I.
\]

In the simple case wherein there is only one set of primary reasons and the agent acts upon them, e.g. slakes his thirst with a glass of water, $pf$-judgement, unconditional judgement and action all converge. In acting as he does, A judges that his primary reasons constitute sufficient grounds to act upon and acts. In the more complex cases where different reasons must be weighed, these reasons compete in the evaluation of the evidence in producing a $pf$-verdict that all things considered, a is better than b.

The schema explains the formulation of an intention as the forming of an unconditional judgement, a judgement conditioned by the whole web of $A^s$
beliefs, affective attitudes, values and information.\textsuperscript{18} Although conditioned by it, it is not conditional upon it, neither is it directed as a desire is to \textit{kinds} of actions (which necessarily include particular actions highly undesirable in other respects).

Given what he believes, desires and values about the future insofar as these are related to his slaking his thirst, A decides that a particular course of action is unconditionally desirable. This is just what it is, according to Davidson, for him to \textit{intend} to slake his thirst.

For A to \textbf{act intentionally}, a transmission of causation from reasons to action through \textit{prima facie} and unconditional judgements must flow uninterrupted. There are three nodes at which it can break down - in passing from a set of primary reasons to a \textit{prima facie} judgement or from the latter to an unconditional judgement or from that last to an action. The troubled politician may have the evidence provided by his beliefs, desires, values and information to infer that it is more desirable to permit US bases than prohibit them, yet he may fail to form that \textit{pf-judgement}. Even with that \textit{pf-judgement} and the realisation that all the available evidence bears it out, he may fail akratically to form the appropriate intention; and even with the appropriate intention he may irrationally fail to act upon it.\textsuperscript{19}

\textsuperscript{18} As Bratman notes (\textit{loc.cit}), the original ontological hope of 'Actions, Reasons and Causes' to reduce intentions to belief-desire pairs has now been abandoned, there now is an element common to all cases of intentionally, viz. acceptance of an all-out evaluation that \textit{\$9} is desirable.

\textsuperscript{19} Not only does the causal sequence not always flow uninterrupted from reasons to action, it can also be completed in a deviant manner, Davidson contends, either by externally or internally wayward causal chains. These deviant chains present an obvious problem to the causal theorist. For the only clear way to rule them out invokes the very concepts the causal theory is supposed to explain - thus Davidson's nervous mountaineer did not let go \textit{intentionally} or did not \textit{try} to let go then and there and so forth. Davidson takes the existence of these wayward chains to destroy the possibility of a causal \textit{analysis} of intentional action.
It is ultimately because his schema tolerates causal anomalies that Davidson insists that there can be no psychophysical laws.²⁰

However, this is not to deny that when an agent acts upon the basis of his reasons a certain causal pathway from reasons to action is in operation. Indeed, Davidson appeals to the causal relation both to bridge the gaps in the schema for intentional action and to rule out deviant causal chains. This is why he says that:

'Unavoidable mention of causality is a cloak for ignorance; we must appeal to the notion of cause when we lack detailed and accurate laws. In the analysis of action, mention of causality takes up some of the slack between analysis and science.'²¹

²⁰ I should say at this point that I am not entirely convinced by Davidson's representation of the causal antecedents of action. David Pears has cast doubt on the identification of the unconditional judgement that Bab with the intention to do a rather than b in 'Intentions as Judgements' pp.222-237 in Zak van Stralen (ed) Philosophical Subjects: Essays presented to P. F. Strawson, Clarendon Press, Oxford, 1980. Pears points out that an intention unlike an unconditional value judgement is effective in that it engages the will - it thus seems perfectly feasible that an agent should judge unconditionally that Bab yet not come to have the intention to do a. Davidson would reject this intuition since he holds that if an agent judges that it would be better to do a rather than b, he must in some sense want to do a more than he wants to do b. On Pears' model, one can form an intention directly on the basis of some desire. We do not need to attempt to decide this dispute here since it is peripheral to the question of the truth of the anomalism of the mental - indeed, if intentions and unconditional judgements can come apart, as Pears suggests, things are even worse for the proponent of psychophysical laws since there is a further juncture at which the causal pathway can be interrupted. If Pears is also right in thinking that intentions are sometimes formed upon the basis of desires alone, the causal schema must be modified to this:

\[ D_1, D_2, \ldots, D_m, \ldots \]
\[ B_1, B_2, \ldots, B_n \]
\[ \vdots \]
\[ C \]
\[ U \]
\[ I \]
\[ A \]

Here U is the unconditional judgement and I the intention. On this account, the akrates moves between U and I irrationally rather than between C and U (equivalent to I) as on Davidson's account.

²¹ 'Freedom to Act' p. 80 in Actions and Events.
An action is performed with a certain intention when it is caused in the
right way by the beliefs and desires that rationalise it.
'What I despair of spelling out is the way in which attitudes must cause actions if they are to
rationalize the action'.

We are now in a better position to understand Davidson's Argument from
Normativeness.

Consider a putative psychological law such as \((L)\) below:
\[
(\forall x)(\forall p)(\forall z)(x \text{ desires that } p \text{ and believes that if he } p^* \\
\text{ then } p \text{ and conditions } z \text{ obtain, then } x \text{ acts }) \quad (L)
\]

Davidson's idea is that laws of the form \((L)\) need not be, as they might first
appear to be, covering laws in a deductive-nomological explanation of
intentional action. For if they are interpreted as covering laws they are, as
the Schema shows, either straightforwardly false or, since unable to exhaust
all ceteris paribus clauses, vacuous.

What accounts for the fact that we find such generalisations compelling and
are reluctant to relinquish them in the face of purported counter-instances is
that they represent normative assumptions about rational action, that is
actions performed for a reason.

---

22 loc. cit p. 79.

23 It is useful here to contrast Davidson's view with Christopher Peacocke's. Peacocke argues in
principles of \((L)\)'s general form do indeed purport to be genuine covering laws but since they are
apriori constitutive principles of rational action, they need a posteriori elements in order to be
genuinely empirical; hence the 'beliefs', 'desires' and 'intentions' that appear in their formulation
must just be place-holders for physical states of some sort - brain states, presumably. Davidson,
on the other hand, would simply deny that \((L)\) really is a covering law.
Thus the conjecture that there is a strict law binding attitudes to behaviour irrespective of the generality of that assumption or its conditionality upon ceteris paribus disclaimers is tantamount to believing that the agent is constrained to act rationally i.e. upon his reasons.\textsuperscript{24}

But there is no more reason to suppose that an agent who has reasons for acting will invariably act upon them than to suppose that a person who has a given set of beliefs and attitudes will invariably draw the conclusion they entail.

As the schema makes manifest, there are distinguishable nodes at which the causal pathway may break down. The Aristotleian syllogism, a special instance of the covering law model, is a limiting case for the general schema in which pf-judgement, unconditional judgement and action coalesce. We might say that an ‘intentional causal pathway’ is operative in these cases.

Davidson writes:

'It is an error to compare a truism like “If a man wants to eat an acorn omelette then he generally will if the opportunity exists and no other desire overrides” with a law that says how fast a body will fall in a vacuum. It is an error because in the latter case, but not the former, we can tell in advance whether the condition holds and we know what allowance to make if it does not. What is needed in the case of action if we are to predict on the basis of beliefs and desires is a quantitative calculus that brings all relevant beliefs and desires into the picture.'\textsuperscript{25}

\textsuperscript{24} Cf. Davidson’s remark at ‘Actions, Reasons, and Causes’ p. 14 in Actions and Events:

‘Suppose that to say a man wanted to turn on the light meant that he would perform any action he believed would accomplish this end. Then the statement of his primary reason for flipping the switch would entail that he flipped the switch – “straightway he acts”, as Aristotle says. In this case there certainly would be a logical connection between reason and action.’

\textsuperscript{25} ‘Psychology as Philosophy’ p. 233 in Actions and Events.
We cannot tell whether the condition holds because we do not know all the relevant beliefs and desires which have contributed to the pf-judgement, if it is one, that eating an acorn omelette is desirable. We do not know if that desire has been reached in the light of specific evidence or whether it is just a vague hankering that the agent finds himself enigmatically with - we know neither the status of that desire nor its proximity to realisation in action nor whether it is differentially sensitive to evidence or information. These are the very sorts of things that we might expect a quantitative calculus bringing in all relevant attitudes to give us.

In the absence of such a calculus, we have no basis for predicting what, given his attitudes, the agent is likely to do. Yet we do manage to ascribe attitudes to agents in ways that they themselves acknowledge as correct. What enables us to do this is the crucial apriori assumption that the behaviour surrounding a certain piece of behaviour is intentional. That is, we assume that certain intentional causal pathways are operative.

It is for this reason that Davidson claims that we necessarily impose certain conditions of coherence, rationality etc in a way which has no echo in physical theory. In physical theory, we do not first make up our minds that certain causal relations hold between phenomena prior to investigation. The task of the physicist is to find out which relations hold.

After making his remark on the difference between the acorn omelette truism and a physical law, Davidson goes on to say that rationality is built into the explanation of acting on a reason in two distinct ways: the first way is, he claims, transparent - a reason is a rational cause and the cause must be a belief and desire in the light of which the action is reasonable. But there is
another less obvious way it enters in - the belief and desire must cause the action in the right way i.e. follow an intentional causal path. He continues:

'... explanation by reasons avoids coping with the complexity of causal factors by singling out one, something it is able to do by omitting to provide within the theory a clear test of when the antecedent condition holds.'

The cost of this mode of explanation, according to Davidson, is that psychological explanation cannot issue in strict laws.

Does this argument give us a compelling reason to believe that there can be no psychophysical or psychological laws? It makes clearer, I think, the part the notion of rationality plays in Davidson's argument. Rationality as a relation between mental states is evinced whenever an intentional causal path is operative. Rationality as an inescapable constraint governing our attribution of attitudes to S is forced upon us to the extent that we identify bits of behaviour as actions.

The cognitive scientist or neuroscientist will, if Davidson is right, perforce read rationality into the behavioural responses of his subjects whenever he judges those responses to be intentional - the subject's responses do not come readily carved up, as it were, with respect to the intentional and the non-intentional (the latter involving, inter alia, perceptual or verbal or interpretative errors). The attempt to entirely eschew all consideration of what the beliefs and desires of the subject might be, will, if Davidson is correct, certainly be hopeless - whatever the mental properties a mature cognitive science posits, our access to them will be mediated by conjectures about the subject's propositional attitudes.

26 op.cit.
27 It is thus incorrect to charge Davidson with conflating the two.
But none of this seems to me to rule out the possibility of psychophysical or psychological laws framed in terms of the posits of such a science.

Mark Johnston has nicely characterised the differences between a 'naive' view of the mental on which it is the goal of a scientific psychology to uncover laws governing those natural mental properties which ultimately explain an agent's behaviour and Davidson's 'interpretative' view according to which:

'... what makes something a mental event or state is not its intrinsic features but rather the fact that it stands in a complex causal web having at its fringes sensory and behavioural events, a causal web upon which an interpretative theory can be slapped.'

where, an 'interpretative theory' is:

'... a consistent and relatively complete set of attitude and action attributions developed under the general guiding principle of making sense of the subject of the theory.'

Johnston argues that such a theory ought to be rejected insofar as we believe that mental states are causally efficacious. I think that he is right.

29 loc. cit. p.424.
§1.2 The Definability of Truth

§1.2.1 Frege's Argument

Davidson wishes to take truth as primitive and use it to illuminate meaning. This does not require him to regard truth as absolutely indefinable, but neither does it prevent him from so doing. I think that Frege has given us good reason why we should take truth to be unanalyzable. It is to his argument that I now turn in an effort to gain clarity about the nature of truth. Frege wrote:

'But could we not maintain that there is truth where there is correspondence in a certain respect? But which respect? For in that case what ought we to do so as to decide whether something is true? We should have to inquire whether it is true that an idea and a reality, say, correspond in some specified respect. And then we should be confronted by a question of the same kind and the game could begin again. So the attempted explanation of truth as correspondence breaks down. And any other attempt to define truth also breaks down. For in a definition certain characteristics would have to be specified. And in application to any particular case, the question would arise whether it were true that the characteristics were present. So we should be going round in a circle. So it seems likely that the content of the word "true" is sui generis and indefinable.' (Thoughts' pp.3-4 in Geach, P (ed) Logical Investigations, Oxford, Basil Blackwell, 1977, pp.1-30.)

Now there is a prima facie worry with this argument: it looks as if Frege's insertion of the phrase 'it is true that' in the fourth sentence 'We should have to inquire whether it is true that ... ' is otiose - it is simply put in to create the illusion of a regress. In actual fact, Frege does provide a reason for its
inclusion, as we shall see. But if we ignore the preceding context of his remarks, we might very well suspect foul play.

Thus Dummett writes in this vein:

'This argument gives a first impression of sophistry. For, one might say, by this means we could show that the notion of truth had to be rejected altogether, whether defined or not: for the same infinite regress could always be generated. Suppose that I wish to find out whether Goldbach's Conjecture is true. Then I must enquire into the truth of the statement "The statement 'Goldbach's Conjecture is true' is true" and so on ... (A). The possibility of the regress thus has nothing to do with whether truth is definable or not.'

Because I will be referring to it in what follows, I have labelled Dummett's crucial claim (A). Dummett's representation of Frege's argument is not perspicuous but I think he means to be saying that since Frege held p and "p" is true to be synonymous, he thought that any enquiry into the truth of p must a fortiori be an enquiry into the truth of the statement "p" is true', which by the equivalence must be an enquiry into the truth of the statement 'the statement "p" is true' is true' and so on. Frege's regress is then a verificatory one - an infinite series of verificatory tasks have to be completed before deciding the truth of any statement.

If Frege's regress really is the verificatory one Dummett takes it to be, his argument for it would seem spurious. For suppose Goldbach is a reporter for the New York Times who has recently been dispatched to Australia. He has

---

2 Frege Philosophy of Language. Duckworth, 1973, (henceforth Frege)p.443, emphasis mine. Dummett puts this forward as part of an objection which he claims refutes Frege's argument for the indefinability of truth without refuting his argument against the kind of definition of truth which the correspondence theory postulates. In 'What in the World could correspond to Truth?' (unpublished), I argued that this was an untenable appraisal of Frege's argument. We can just note here in passing that Dummett's rendition of Frege's argument cannot be exactly right because in the Fregean original truth is predicated of thoughts not sentences or statements. This might not seem a very significant point but it does lead Dummett to misconstrue the character of Frege's argument.
never seen the Prime Minister of Australia but has some indirect evidence for thinking that he must be a small man. He makes the following conjecture: 'Hawke is small'.

Now if (A) were correct, Goldbach, in wanting to find out whether his own conjecture were true, would have to first settle the truth of the statement 'the statement "Goldbach's conjecture is true" is true' and so forth. But this is patent nonsense. All Goldbach has to do to decide whether his conjecture is true is to see whether Hawke is small ... charity alone suggests a different interpretation of Frege's argument.

Frege's criticism is directed, in the first instance, at the correspondence theory of truth which asserts that truth is a relation between discrete linguistic (or mental) items (sentences, statements, thoughts, utterances, beliefs etc) and discrete bits of reality (facts, states of affairs etc.). The correspondence theorist traditionally claimed: 'It is because a thought (or statement) corresponds to a fact (or state of affairs) that it is true.'

What Frege is puzzled by is the nature of this correspondence. He points out in the text preceding (F1) that the correspondence in question between the thought and the fact cannot be coincidence in all respects (i.e. identity) since the relata are distinct entities. It must be a correspondence in a certain specific respect. A picture of Cologne Cathedral (or a photograph of it) might offer a recognisable representation of the front elevation of the cathedral. However, and this is the crucial point, in determining whether the relation of representation holds between this picture and the cathedral, we're enquiring, Frege contends, as to whether a certain judgement is true - namely the judgement that the picture is a representation of Cologne Cathedral.
What Frege means here is not that it is an evidential prerequisite for settling the truth of the claim that the picture corresponds to Cologne Cathedral in a certain respect that we first settle the truth of the judgement that it is true that the correspondence holds in that respect. His point is, rather, that this is what making the original judgement consists in.

This is, I think, Frege's crucial premise. Accordingly, I label it (CP):

\[(\forall \text{ thoughts } p)(\forall \text{ states of affairs } \Sigma)(\text{to judge that } p \text{ and } \Sigma \text{ correspond in respect } X \text{ is to judge that the thought that they so correspond is true}) \quad \text{(CP)}\]

Frege then notes that any further definition will be of the form:

the thought that \( p \) is true iff the thought that \( p \) has certain characteristics, \( F \).

So (CP) can, given a certain plausible assumption about the nature of judgement which we shall examine later, be generalised to:

\[(\forall \text{ thoughts } p)(\forall \text{ properties } F)(\text{to judge that } p \text{ has the property } F \text{ is to judge that the thought that } p \text{ has } F \text{ is true}) \quad \text{(CP*)}\]

Given (CP) (and, more generally (CP*)), Frege goes on to argue that if the truth of thoughts consisted in a correspondence relation such as that which holds between picture and object depicted, then, insofar as we judge that the correspondence obtains between any given thought \( p \) and a certain state of affairs \( \Sigma \), we judge that the thought that the correspondence obtains in the relevant respect \( X \) is true. But judging that that latter thought is true is a matter of judging that a yet higher order corresponding thought is true which is a matter of ... ad infinitum.
It should now be clear that Dummett's interpretation of Frege's argument is mistaken. Frege is not committed to the claim at (A) which Dummett attributes to him. For it only follows from Frege's argument that one must decide whether the metalinguistic statement "\(p\) is true" is itself true before deciding the truth of \(p\) if every such judgement involved establishing that a correspondence holds between \(p\) and some state of affairs \(\Sigma\) in a certain respect \(X\) ... but this is precisely what Frege is denying!

Even if Dummett has misinterpreted Frege here, a further argument of his might be thought to enable us to stop Frege's regress: "It is true enough that in determining that some statement \(A\) is true, I thereby determine the truth of infinitely many other statements, namely '\(A\) is true'; 'the statement '\(A\) is true' is true' etc. But there is no harm in this, as long as we recognise that the truth of every statement in this series is determined simultaneously: the regress would be vicious only if it were supposed that in order to determine the truth of any member of the series, I had first to determine that of the next term."

Dummett's claim is that in determining the truth of \(A\), I thereby determine the truth of its infinitely many 'higher order truth-evaluators', as one might call them: '\(A\) is true'; 'the statement '\(A\) is true' is true' etc. This is surely right for Dummett's verificatory regress as the case of Goldbach the reporter shows: by simply observing that Hawke is small, Goldbach settles the truth of his conjecture that Hawke is small, and in the process indirectly verifies all logical consequences of the conjecture amongst which are 'Goldbach's conjecture is true'; 'the statement "Goldbach's conjecture is true" is true', etc. Might it not be that the Fregean regress can be blocked in a similar way to Dummett's?

---

3 Frege p.443.
Dummett's regress is a regress of enquiry - an infinite series of verificatory tasks confronts the would-be verifier of the truth of any statement A. Frege's regress, in contrast, is not a regress of enquiry, but an explanatory regress. What is responsible for generating that regress in the first instance is Frege's principle (CP*) - that to judge that p just is to judge that the thought that p is true.

The regress arises in the following manner: the proponent of a general truth definition or theory of truth puts forward his general definition of truth - "p" is true iff p has F, where p is, in Fregean terminology, a thought. To Frege's query 'what does the truth of the thought that p consist in?', the truth-definer replies 'the thought that p has F.' Frege then observes that by (CP*), to judge that the thought that p has F is to judge that the thought that p has F is true. So at this stage, we have that for the thought that p to be true is for the thought that p has F to be true.

'But', Frege objects, 'this definition is clearly circular since "true" occurs in the definiens'. So if the truth-definer is not to be saddled with a circular explanation of the meaning of his definiendum, he must try to eliminate the occurrence of the definiendum from the definiens. This involves replacing 'is true' by 'has F'.

So we get: for p to be true is for the thought that p has F to be true which is for the thought that p has F to itself have F. But by (CP*) and the truth definition, to judge that the thought that p has F has F is to judge that the thought that p has F has F itself has F ... and so on ad infinitum.

Thus, the truth-theorist's definition is either flagrantly circular or leads to a vicious infinite regress of explanations. Hence, Frege concludes, any attempt, and not just the correspondence theorist's attempt, to define truth or to explain what it consists in breaks down.
It is clear that we cannot answer Frege’s argument by invoking Dummett’s point that it is possible to verify that it is true that \( p \) without first having to verify any higher order truth-evaluator for \( p \). Certainly this is possible but it has nothing to do with Frege’s worry.

If we are to rebut Frege’s argument, we must attack \((CP^*)\), the principle that states that to judge that a thought has the property \( F \) which a theory of truth ascribes to it as the condition for it to be true just is to judge that the thought that \( p \) has \( F \) is true.

An argument against this principle might proceed as follows:

At least for suitably restricted \( p \), the proposition or thought that the thought that \( p \) has \( F \) is true entails and is entailed by the thought that \( p \) has \( F \). Error results when we use this fact to argue that the judgement that the thought that \( p \) has \( F \) is true just is what the judgement that \( p \) has \( F \) really consists in. For, in order for these distinct judgements to be equivalent, their contents must be equivalent and not just co-extensive - even necessary co-extensiveness is too weak. But, the objection continues, no thought to the effect that another thought is true can possibly be equivalent in meaning to that latter thought.

Does the thought that \( p \) have the same content as the thought that \( p \) is true? Is the sentence \( A \) synonymous with the sentence \( 'A' \ is true' \)? My intuitions about synonymy unfortunately are not fine enough to answer these questions.

Fortunately, we do not need to try to decide them. Frege’s argument does not make use of any alleged synonymy between \( 'p' \) and "\( p \) is true". \((CP^*)\) suggests that judging that the thought that \( p \) has \( F \) is true is what judging
that \( p \) has \( F \) consists in. What does it mean to say that the former judgement is what the latter 'consists in'? Just this - that any adequate theoretical account of what it is to judge that \( p \) has \( F \) would proceed in terms of an account of judging that the thought that \( p \) has \( F \) is true. The latter type of judgement will be used, Frege thinks, to give an acceptable theoretical explanation of what the former type of judgement really is in exactly the same way as the firing of C-fibres will be used in an acceptable theoretical explanation, according to the physicalist, of what having a sensation of pain is. No one would claim that 'my C-fibres are firing' is equivalent in meaning to 'I have a toothache', yet the physicalist's thesis is that the latter statement can be reduced to the former, or, in our terminology, having one's C-fibres fire is what having a toothache really consists in. Questions of synonymy simply do not obtrude (or ought not to) when our interest lies in theoretical explanation or reduction.

I doubt that a convincing case can be made for \( (CP^*) \) on analytic grounds. For the fact is that it is not at all clear that 'A' and 'A is true' are synonymous.\(^4 \) Yet if \( (CP^*) \) is to be interpreted as asserting the identity of two apparently distinct judgements - to wit: the judgement that \( p \) has the property \( F \) and the judgement that the thought that \( p \) has \( F \) is true - it requires that the contents of those judgements be equivalent, which we have reasons for doubting.

The 'analyticity' advocate for Frege might be tempted to weaken the synonymy requirement in \( (CP^*) \) to necessary co-extensiveness: Frege's regress might still go through if judging that the thought that \( p \) is true is necessarily co-extensive with judging that \( p \).

\(^4 \) And it is even less clear how one can possibly decide!
Let us see whether this is correct. Consider a parallel case. Suppose A wishes to know what 'bachelor' means. B obliges A by producing the standard definition: 'x is a bachelor iff x is an unmarried man'. Wishing to make life hard for B, A informs him that he understands perfectly well what 'unmarried man' means but that he is worried that B's definition is circular. 'How so?', asks B. A replies calmly: Your definition is logically equivalent to "x is a bachelor iff x is an unmarried man and either Frege was a bachelor or Frege was not a bachelor", which, since it uses the definiendum in the definiens, is straightforwardly circular.

It is clear that A's tactics are unfair. He has simply deployed a logical equivalence in order to spoil a perfectly acceptable definition. The truth-definer's objection to Frege's regress argument is that Frege has done precisely the same thing - he has used the undoubted logical equivalence between 'p' and "'p' is true" to insinuate a circularity in an otherwise acceptable truth-definition of the form: "'p' is true iff p has F". The same tactics of artificial definition-spoiling are at work in Frege's argument, the truth-definer will claim. What conceivably could rule out unwanted logical equivalences and legitimately warrant the infinite series of iterated demands that Frege makes upon the truth-definer to say how, at each turn, the truth of the higher order truth-evaluator of any thought p is to be defined, is a synonymy between 'p' and "'p' is true". But this is precisely what is in doubt.

So the truth-definer's objection to Frege's procedure is not that (CP) or (CP*) is false - qua principles concerning the logical equivalence of 'p' and 'p is true' they are unexceptionable - his worry has to do with Frege's entitlement.
to use the equivalence to generate a regress, given that it is possible to spoil any definition by choosing a suitably perverse set of logical equivalences.

This objection strikes me as a forceful one. Frege would have been unwise to invest so much in a putative synonymy between 'p' and "'p" is true" and even if he had, it is unclear that the return on this investment would have included a vindication of his regress argument. I do not believe he did so.

What is suggested by his remarks is the thesis adverted to earlier: that when we attempt to give a theoretical account of what an agent is doing in judging that p, we shall have to explain his mental activity as precisely one of evaluating the truth of the thought that p, on pain of not being able to explain it at all. Not only his judgements but all of his contentful psychological states permit of a smooth and powerful initial explanation in terms of attitudes that an agent has to the truth of the propositions which express their contents.

Thus we can make a start upon a theory of desire by characterising the 'desideratory force' of a desire that p as one of desiring that the thought that p be true, of an intention that p's 'intentional force' as one of intending to bring it about that p be true, and of a judgement that p's 'judicative force' as that of judging that p is true.

Assuredly, this is only the start of such a theory, but it is an important start for it allows us to explain the transitions from cognitive state to cognitive state as deductive transitions between states of differing modalities but common contents. Even if the theory has to be complicated to allow for irrational but still intentional action in the manner suggested by Davidson's decision theory as in §1.1.2, it remains that if we conceive of propositional attitudes as attitudes to the truth of propositions, we can explain how the
ratiocinations of an agent can lead on to actions or the formation of intentions and long term plans.

When \((CP^*)\) is interpreted along such causal-explanatory lines, the previous objection to it no longer applies. It would be a travesty to allege that Frege is yet still making use of a gerrymandered logical equivalence just to generate a regress if he reasoned as above. For the fact that 'A' and 'A is true' are logically equivalent does not feature at all in our theoretical reasons for explaining the cognitive and conative modalities of states with a common content \(p\) in terms of attitudes toward the truth of \(p\). Our grounds for adopting this explanation were quite independent of such considerations, having to do with the possibility of explaining causal transitions between such mental states in terms of the deductive relations between states sharing a common content but differing in the agent's attitude to the truth of that content.

Such a theory does seem to me to provide a good reason for accepting \((CP^*)\) and thus for believing that Frege's argument to the conclusion that truth is sui generis and indefinable is sound.
§ 1.2.2 Frege and Tarski

Now it might be thought that the Frege Argument, as I shall call it, is quite inconsistent with Tarski's formally rigorous means of defining truth for any given language. Hasn't Tarski simply refuted Frege? I do not believe he has but there are two challenges any defender of the Frege Argument needs to face.

He first of all has to show that Tarski's enterprise is consistent with the conclusion of the Frege Argument; he then has to show how Tarski's definition of truth, insofar as it defines truth for L via the primitive concept of satisfaction, is not (at least covertly) a Correspondence Theory of Truth since satisfaction is held to be an inherently relational notion, relating linguistic expressions to objects (or sequences thereof) in the world.

One philosopher who takes a very deflationary attitude to the philosophical illumination afforded by Tarskian truth definitions is Hilary Putnam. He offers the following example to support his view:

(i) Let 'snow-shaped' be a predicate applying to all and only those inscriptions having the syntactic shape 'snow is white'.
(ii) Let 'grass-shaped' be a predicate applying to all and only those inscriptions having the syntactic shape 'grass is green'.
(iii) Call a sentence snow-good iff it is the case that:
(a) the sentence in question is snow-shaped; and
(b) snow is white.
(iv) Call a sentence grass-good iff it is the case that:
(a) the sentence in question is grass-shaped; and
(b) grass is green.
(v) Call a sentence L-true iff the sentence is either (a) snow-good or (b) grass-good.

As Davidson in 'True to the Facts' pp. 37-54 in Truth and Interpretation and John Wallace in 'Belief and Satisfaction', Nous vol. 6 (1972) have claimed.
The property of being L-true just defined is the property that would correspond to the truth predicate for a language with just two sentences - 'snow is white' and 'grass is green'. So if we have a metalanguage which contains the names and translations of just these two sentences (which comprise the total stock of wffs for the object language) and which also has sufficient syntactic resources for structural description (devices for spelling and talking about the shapes of sentences) and has quantifiers ranging over sentences and standard truth-functions, then in such a language, we can easily define the predicates:

1. 'snow-shaped';
2. 'grass-shaped';
3. 'snow-good';
4. 'grass-good';
5. L-true.

The language is strong enough to give a Tarskian truth definition and we can thus prove the following two theorems assuming standard axioms for logic and syntax:

(A) 'snow is white' is L-true iff snow is white;
(B) 'grass is green' is L-true iff grass is green.

With this example in mind, Putnam declares:
'What I deny is that the so-called "truth-predicate" that Tarski defines is in any way similar in meaning to the intuitive predicate "is true"'  

The reason he gives for this bold claim is that the property of being L-true as defined for our two-sentence language above is defined without reference to speakers or their use of words since whether or not an arbitrary string of letters is L-true depends upon just three things:

(1) How the string is composed;
(2) Whether snow is white;
(3) Whether grass is green.

---

The string of letters 'snow is white' has the property of being L-true in all worlds in which snow is white, even ones wherein it means 'water is a liquid' and since it is easy to imagine counterfactual situations in which a sentence could be L-true and yet not be true, Putnam claims, the property of being true has not been conceptually analysed by the predicate L-true.

Now whilst I think that Putnam's conclusion is correct, I find his denial that the Tarskian predicate is in any way similar in meaning to the intuitive truth-predicate somewhat startling. One who endorses the Frege Argument need not be committed to this extreme claim - he simply denies that the Tarskian predicate can be understood independently of any intuitive grasp of the concept of truth.

Putnam's trivialisation argument does highlight the need to relate a Tarskian definition of truth to the practice of speaking a language in order to gain conceptual insight into truth. Michael Dummett has forcefully argued for just this conclusion. Dummett observes that while Tarski defined the sentential predicate 'is true' for a particular language, he did not and in fact proved one could not, at least by his procedure, define truth in general. Hence, according to Dummett, Tarski was unable to say what made each of the individual truth definitions a definition of truth - the property of truth had been

\[7\text{Frege pp.295–298; compare also 'Truth' in Truth and Other Enigmas pp.2–4, 8; 'Postscript' (to 'Truth') pp.19–20; 'Frege's Distinction between Sense and Reference' pp.123–124; Preface to Truth and Other Enigmas pp.xx–xxi.}\]
specified disjunctively where we sought an essential characteristic which each of the separate truth predicates share.\textsuperscript{3}

Dummett gives the instructive analogy of winning in a game - truth, like winning, is a property we value and seek to secure in our semantic transactions with others and with the world. Suppose we did not know what winning at a game was. We wouldn't be much the wiser if we were just told the definition of winning for each of several different sorts of games for we would still not have been shown what makes each situation a situation of winning. Precisely the same holds for truth. Were we to encounter speakers of a foreign language for which we were given a truth definition, we would be in the dark about what predicate this definition explicated until we discovered for ourselves how the truth definition actually applied to this language, what features of the natives' linguistic behaviour were meant to be systematically described by the predicate etc.

Dummett's attitude strikes me as a more temperate one - he does not have to deny that the Tarskian predicate is similar in meaning to the intuitive truth predicate, he simply denies that a Tarskian truth-definition by itself offers us much insight into the concept of truth.

Tarski showed how to define a predicate which is co-extensive with the intuitive truth predicate as that is relativised to the sentences of a particular interpreted formal language. With that much Putnam agrees. Putnam notes

\textsuperscript{8} Compare his claim at Frege p.455:

"An account, however accurate, of the conditions under which some predicate is rightly applied may thus miss important intuitive features of its meaning; in particular, it may leave out what we take to be the point of our use of the predicate. A philosophical account of the notion of truth can thus not necessarily be attained by a definition of the predicate "true", even if one is possible, since such a definition may be correct only in the sense that it specifies correctly the application of the predicate, while leaving the connections between this predicate and other notions quite obscure."
that a supporter of the philosophical riches of Tarskian truth definitions will look to Tarski's criterion of correctness for truth-definitions, Convention(T) for insight into what truth in general is. But, Putnam claims, there is a problem in deploying Con(T) for this purpose.

To read the principle as saying that all instances of the following schema must be true:

\[(T) \text{ } s \text{ is true-in-} L \iff p^9\]

is obviously circular if our interest in using the principle is to find out what truth in general is, because we are making use of an intuitive grasp of truth: 'If we know that all instances of the above schema are true, then of course we know the predicate "L-true" is coextensive with the predicate "true in L". To say this may be to say something informative about the notion of truth; but since to say this we have to use the notion of truth (in the language ML) we have not really successfully answered any of the deep philosophical worries about what truth really comes to.'

Putnam notes that what Tarski required was not that all instances of \((T)\) be true but that they should all be *theorems* of the metalanguage ML: 'But this will not do either; if the language ML is \(\omega\)-inconsistent, then it may be the case that some instances of the above schema are theorems of ML although (in the metametalanguage) one can easily see that they are not all (in the intuitive sense) true.'

I am not entirely sure what Putnam has in mind in this passage in his talk of an \(\omega\)-inconsistent metalanguage. There is an initial problem with the passage - the 'some instances' in the second line should be 'all instances', since otherwise there is no inconsistency. The general point that we should only be able to prove true T-sentences is obviously right. An \(\omega\)-inconsistent language is a notion that requires some explication as it is theories and not languages which are usually said to be \(\omega\)-inconsistent. A first order theory \(K\) with a denumerable number of sentences is \(\omega\)-inconsistent according to

---

9 Use underlining to denote the structural-descriptive name in the metalanguage for the object language sentence \(s\) translated by \(p\) in the metalanguage.

10 *loc. cit* p.41.

11 *loc. cit.*
Mendelson at p.142 of his Mathematical Logic if and only if, for every wff $A(x)$ of $K$, if $\vdash_k A(n)$ for every natural number $n$, then it is not the case that $\vdash_k (\exists x) \neg A(x)$. Thus stated, there is no clear application of this notion to the case at hand, since the only semantic property which could substitute for our $A$—viz. truth-in-$L$—is precisely not a constant property of $L$-sentences, since not all $L$-sentences, in the general case, are true!

There is nothing circular or illegitimate about requiring instances of the $(T)$ schema be provable for the purposes of constructing a Tarskian truth definition for a (suitably regimented) interpreted object language using an ($\omega$-consistent!) semantic theory in one's metalanguage, it is just that the definition by itself will not impart an understanding of the concept of truth. To do that we need at least to relate the concept of truth to the actual practice of speaking a language.

Even if one thought that Tarski's definition told us the whole truth about truth, and that the notion of satisfaction (in terms of which Tarski defined truth for the sentences of a given OL) is conceptually more basic than that of truth, we should note that there is nothing obligatory about defining truth for the OL in Tarski's way.

Frege had already indicated an alternative method. ¹² The reason Tarski had to introduce the auxiliary notion of satisfaction to define truth was that logical connectives can stand between predicates as well as sentences—as in

---

'Some philosophers are long-winded and boring'. On pain of admitting equivocal readings of the connectives as they occur in sentential and predicative contexts, we have to either explain truth in terms of satisfaction or satisfaction in terms of truth. Tarski opted for the former method. But it is possible and, I believe, intuitively more plausible, to reverse this direction. The Fregean strategy takes the sentence-forming role of connectives as basic. We then define the relation of satisfaction which holds between an expression and its object(s) in terms of the truth-value of a sentence which results from substituting a singular term referring to that (those) object(s) in that predicate. If the language contains no name for the object(s), we explain the satisfaction relation in terms of the truth-value in some finite extension of the language which does contain a proper name for that (those) object(s) of a sentence which results when a singular term which refers to that (those) object(s) upon extending the language is substituted in the predicate. It is assumed that for any object there is an extension of the language containing a singular term for that object. At no stage is it assumed that there is an extension of the language containing a singular term for every object. So, we explain the satisfaction conditions for the complex predicate 'is long-winded and boring' thus: 

\( y \) satisfies 'is long-winded and boring' iff upon that extension of the language on which \( \xi \) denotes \( y \), '\( \xi \) is long-winded' is true and '\( \xi \) is boring' is true.

Even if the compound thesis that Tarskian truth-definitions present us with the whole explanation of truth and that such truth-definitions amount to versions of the Correspondence Theory of Truth to the extent that they

\[\text{13 Cf. Evans loc. cit.}\]
define truth by the relational notion of satisfaction were right, Fregean truth definitions present us with an alternative means to the same end. The Frege Argument provides us with a reason for believing that this alternative might better capture our fundamental intuitions about the centrality of truth in the explanation of cognitive and semantic notions.
§1.3 TRUTH THEORIES AND THEORIES OF MEANING.

§1.3.1 Truth and Meaning

What should a theory of meaning do? Primarily, effect a transition from physics to semantics - that is, license the redescriptions of certain issuings of vocables as linguistic acts of a certain kind and with a certain content. Correlatively, what distinguishes Charles from monoglot julio is this ability to successfully perform such reconstruals of English utterances. Davidson's question is what julio would have to know to acquire this ability. If we could explicitly state the knowledge which would suffice for the interpretation of English utterances in the form of a theory, then julio, by knowing the theory, could acquire the ability ... (0).

This at any rate is Davidson's (plausible) starting point. Accordingly, I label it (0). But we need to lay down more specific requirements on theories of meaning in order to advance matters. Davidson emphasises the following two basic conditions:

\( (C_1) \) A theory of meaning must assign a meaning to all the expressions of the object language \( L \).

\( (C_2) \) A theory of meaning must exhibit the meaning of complex expressions as a function of the meanings of a finite number of semantic primitives compounded by a finite number of compositional devices.

\( (C_1) \) is uncontroversial - a theory of meaning which failed to do this could not possibly pass the adequacy constraint in (0). In contrast, \( (C_2) \) is highly controversial. I find the controversy surprising. The constraint is a semantic

---

1 As Ernest Lepore has observed ('In Defense of Davidson: Linguistics and Philosophy vol.5 no.2 1982, pp.275-294), the requirement that we explicitly state this knowledge is not idle since: "it is impossible to know that someone knows that something is the case without sharing this knowledge ... if I know that you know the \( \emptyset \), then there is no guarantee that I know the \( \emptyset \) ... if I know that you know that 'snow is white' is true if and only if snow is white, then I know that 'snow is white' is true if and only if snow is white. But if I know that you know the truth-conditions for 'snow is white' it does not follow that I know the truth-conditions for this sentence." p.281.
analogue of Chomsky's claim that an adequate grammar for a language must be recursively based given the empirical facts that humans acquire an indefinite extensible grammatical competence in an extremely short period of time on the basis of exposure to piecemeal bits of (often impoverished) syntactic data. Pari passu for semantics - if a speaker's semantic competence were not recursively generated from finite semantic features, it would be utterly mysterious how the speaker could understand a potential infinity of novel sentences.  

It is Davidson's strong claim that a language containing infinitely many semantic primitives is unlearnable which has caused dissent. But this claim seems very plausible to me. Given that it takes the speaker of L a finite time to learn each of its semantic primitives, a language with non-finitely many primitives could not be acquired in finite time. Again, without a set of recursive rules for generating the meanings of complex expressions from semantically primitive expressions, to learn L would require one to learn infinitely many facts - one corresponding to the meaning of each of the infinitely many complex expressions formulable in L.

---

2 As Gareth Evans ('Semantic Theory and Tacit Knowledge', in Holtzman, S. and Leich, C. Wittgenstein: To Follow a Rule, Routledge and Kegan Paul, 1981), Martin Davies (Meaning, Quantification and Necessity: Themes in Philosophical Logic, Routledge and Kegan Paul, 1981) and William Lycan (Logical Form in Natural Language, Bradford Books, MIT Press, Cambridge, Massachusetts, 1984) have each independently emphasised, the need to find structure in a speaker's semantic competence is generated just from the fact that the speaker can understand novel utterances on the basis of familiarity with the words and constructions as they appear in other sentences. Even if we spoke a finite language, the need for a structure-discerning theory would still hold.

3 'Theories of Meaning and Learnable Languages' and elsewhere in Truth and Interpretation. Here, $\alpha$ is a semantic primitive if it is impossible to come to know the meanings of sentences containing $\alpha$ by inductively projecting from the meanings of sentences not containing $\alpha$. 
So, it would seem right to hold with Davidson that a learnable language is one that deploys only finitely many semantic primitives and finitely many recursive compositional devices.4

There are several other constraints which adequate theories of meaning should meet. Two of the more important are:

(C3) That the theory provide a means of disambiguating sentences.

(C4) That the theory's predictions as to the system of implications between sentences of L largely be borne out by the semantic intuitions of competent speakers of L.

Davidson had originally believed that a suitably constrained Tarskian truth-theory which recursively characterised the predicate 'true-in-L' could be used to reconstruct the meanings of sentences from the concept of truth alone. For the theorems of such a theory - the oft-dubbed 'T-sentences' of the form \((T) \varphi \text{ is true-in-} L \leftrightarrow p\) (with '↔' the material biconditional) - use either the object language sentence named by \(\varphi\) itself (in the case where the metalanguage M is an extension of the object language L), or its metalinguistic translation (when M is distinct from L), on the right hand side of the biconditional in \((T)\) to state conditions under which each sentence of the object language is true.

Davidson had thought that we could avoid pairing an object language sentence on the LHS with anything other than its metalinguistic translation on the RHS if we but required:

(1) The T-sentence to be true.

(2) The T-sentence to be a theorem of a finitely axiomatised theory which entailed a true T-sentence for every sentence of the object language L as a result of discerning the semantic structure of those sentences.

4 Correspondingly, it would also seem right to hold with Davidson that, contraposing, semantic theories which entail that a speaker must learn infinitely many rules or meanings are a fortiori inadequate.
In this way, 'deviant' T-sentences such as:

\[(T^*) \text{ 'Snow is white' is true-in-L } \leftrightarrow \text{ water is wet.}\]

were to be ruled out. Moreover, had (1) and (2) sufficed to ensure that all and only meaning-giving T-sentences appeared as theorems of the truth theory, the crucial constraint \((C_1)\) - that a theory of meaning for L state the meanings of all L-expressions - would have been satisfied, and along with it the constraints \((C_2), (C_3)\) and \((C_4)\) would also have been satisfied.

But as John Foster and Brian Loar independently showed, (1) and (2) imposed too coarse a mesh on possible truth theories. To take Foster's example, we can show that a deviant interpretation for the universal quantifier, represented as \(\forall^*\), has the same truth-relevant properties as its counterpart \(\forall\) in L even though \(\forall^*\) is translated as 'For every object ... the earth moves and ... '. As Foster says:

'Given a language \(X\) and its truth-theory \(Y\), a false interpretation of \(X\) is compatible with \(Y\) if revising the meanings of the \(X\) sentences to accord with this interpretation preserves in all cases their grammar and truth-values.'

The problem is clearly caused by the extensionality of the biconditional. Our T-sentences are to be read 'either both s is true-in-L and p or neither s is true-in-L nor p'.

Let us put ourselves in Julio's shoes as he attempts to understand the linguistic interchanges between English-speaking Charles and English-speaking Diana. Our primary evidence for a theory as to what Charles' words mean, according to Davidson, will then be the class of sentences Charles holds true. It is this class together with the states of affairs obtaining in the world (as Julio perceives them) which provide Julio with the empirical evidence.

---


against which his truth-theory is to be tested. But if the connective between
$s$ and $p$ in the $T$-sentences is extensional and our evidence just consists in
the class of sentences Charles holds true together with the facts as we see
them, then no single pairing of sentences with states of affairs can be
favoured over any other extensionally correct one. To single out any one
such mapping as a 'meaning-giving' one is unjustified since all that we can
legitimately require so far is materially and grammatically correct $T$-
sentences representing mappings of held-true object language sentences
onto conditions in the world.

An extensional connective in an instance of (T) simply tells us that, as a
matter of brute empirical fact, either Charles holds 'It is raining' to be true
and it is raining or Charles does not hold 'It is raining' to be true and it is not
raining. Plainly, if he is to interpret Charles' utterance of 'It is raining' when it
is raining as meaning that it is raining, Julio needs to know that Charles
utters this sentence because it is raining. But this piece of knowledge
cannot be imparted by an extensional biconditional, or so it would seem.7

An obvious move is to replace '↔' by a modal biconditional. This strikes me
as a retrograde step theoretically - for where before we made use of a truth-
functional connective whose semantics is transparent, we're now proposing
to use a non-truth-functional connective in its stead whose exact semantics
has not been definitively settled and which might require ontological
commitment to possibilia or somesuch. Quinean scruples aside, a modal
biconditional is heir to the very same difficulties as its extensional

7 Alternatively, if we insist on extensionality, the truth-theory simply becomes a mechanical way
of mapping sentences onto states of affairs, in which case it is at best misleading to interpret the
semantic terms 'true', 'satisfies', etc as meaning what they do - these terms should really be
uninterpreted.
predecessor. For all we need do to admit deviant T-sentences into our truth-theory is reinterpret the modal ‘iff’ as ‘iff ... and 2 + 2 = 4’.

All things considered, modal strengthenings weaken the theory. We should therefore keep our biconditional extensional if at all possible. If we could then somehow pick out the interpretational truth-theory from the merely true truth-theories, we might yet be able to state that knowledge which would suffice to interpret a speaker’s words.

Let us go back to our starting point and try to be a little more precise about the ability Charles possesses which Julio lacks. We described it very generally as an ability to reconstrue Diana’s sequences of vocables as linguistic acts of a certain kind with a certain content.

Suppose Diana asks Charles somewhat truculently ‘Were you born in a tent?’ Charles unhesitatingly reconstrues this rather odd question as a command to shut the door and dutifully obeys. But had Charles not understood Diana’s utterance in the first instance as a question as to whether he had been born in a tent, he would never have sought, in the light of its contextual oddity, to reconstrue it at all - let alone as a command to shut the door.

Following Barry Taylor\(^8\), I will call an initial classification of an utterance in terms of its mode and content ‘primary’ if it does not depend on some other redescriptions of that act. Then we can say that what distinguishes Charles from Julio qua English speaker is the ability to find primary redescriptions of utterances of English sentences as being in a certain mode and with a particular content. A dimmer Charles, though failing to realise that Diana

wanted him to shut the door, could not be accused of *semantic* incompetence if he could produce the primary redescription.

As described, there are two components to Charles' ability to find primary redescriptions of English utterances. Correlatively, a theory of meaning for English should consist of two sub-theories:

(i) *A theory of force* which classifies utterances into ones of a certain mode according to the discernible intentions of the utterer or contextual features of the utterance circumscribed by various tacit conventions English speakers are party to.

(ii) *A theory of sense* which assigns contents to utterances whose mode has been fixed by the theory of force.

Taylor maintains very plausibly that the content assigned to an utterance in its primary redescription depends solely upon the sentence uttered - even though contextual features are germane to the determination of the *mode* of *'La terra si muove'* (which utterance may, in secondary redescription, serve as a confession by Julio of his love for Diana), in primary redescription the utterance can only be an assertion/query/command/... concerning the Earth's moving.

Now the theory of sense, if it is to be adequate, must generate primary redescriptions of utterances of English sentences which accord with those which English speakers themselves are disposed to give.9 Furthermore, just because the content of a primary redescription depends solely upon that of the sentence uttered and because the meaning of that sentence must in turn systematically depend upon the meanings of its parts together with its compositional devices, a theory of sense for L should, according to Taylor10, assume the following general form:

9 For otherwise Julio could not use the theory to acquire the very same ability possessed by Charles.
10 *Modes* ch. 1 pp. 3-5.
(1) It should be an effectively axiomatised theory formulated in a metalanguage M for L with a recursive syntax for L.

(2) The recursive syntax will discern the semantically primitive expressions of L together with the compositional devices of L and for each of the finitely many primitives there will be an axiom of the theory giving the conditions under which the semantic concepts apply to it and to each of the finitely many compositional devices there will correspond a recursion clause specifying how the semantic concepts apply to compound expressions formed from the primitive expressions in accord with that compositional rule.

(3) The theory will contain a distinguished context \( \xi \) and a canonical proof theory \( \varphi \) such that for each indicative sentence \( s \) of L, there is a canonically provable instance of the schema: 

\[
(1) \xi(s, p) \tag{1}
\]

where \( s \) is the standard name for \( s \) in the metalanguage M (e.g. a quotation-mark name of \( s \) or a numeral for a Goedel-number thereof) and \( p \) is a sentence of M that translates \( s \) in M.

The task of the canonical proof theory \( \varphi \) is to display in a rigorous fashion precisely how the contents of sentences of L systematically depend upon those of their parts. The requirement on the proof-theory is:

'... that, in proving such instances of (1), it invokes precisely those semantic axioms which concern syntactic primitives and compositional modes displayed in the S concerned, and invokes them moreover in an order determined by that in which the preferred syntax uses those modes to generates from the given primitives.'

The canonical proof theory thus supplies us with an effective procedure for discerning the semantic structure of each sentence of L. The canonical provability of each instance of (1) can therefore be identified with the theory

---

\( \xi(s, q) \) is an argument-place for the name of a sentence and a sentence. I think Taylor must intend his distinguished context \( \xi \) to be not entirely unstructured for otherwise if we prove that \( p \rightarrow q \) and also \( \xi(s, p) \), we do not know whether we can infer that \( \xi(s, q) \). So we need at least a substitutivity rule licensing this inference.

\( \xi(s, q) \) is an argument-place for the name of a sentence and a sentence. I think Taylor must intend his distinguished context \( \xi \) to be not entirely unstructured for otherwise if we prove that \( p \rightarrow q \) and also \( \xi(s, p) \), we do not know whether we can infer that \( \xi(s, q) \). So we need at least a substitutivity rule licensing this inference.

\( \xi(s, q) \) is an argument-place for the name of a sentence and a sentence. I think Taylor must intend his distinguished context \( \xi \) to be not entirely unstructured for otherwise if we prove that \( p \rightarrow q \) and also \( \xi(s, p) \), we do not know whether we can infer that \( \xi(s, q) \). So we need at least a substitutivity rule licensing this inference.

\( \xi(s, q) \) is an argument-place for the name of a sentence and a sentence. I think Taylor must intend his distinguished context \( \xi \) to be not entirely unstructured for otherwise if we prove that \( p \rightarrow q \) and also \( \xi(s, p) \), we do not know whether we can infer that \( \xi(s, q) \). So we need at least a substitutivity rule licensing this inference.
of sense's way of stating that, in the case of indicative $s$, the content associated with $s$ is that $p$.\textsuperscript{13} Although it is up to the theory of sense to specify its context $\xi$, for the theory to combine successfully with the theory of force, the object language sentence designated by $s$ must be paired with its metalinguistic translation in the instances of the schema $\xi(s,p)$. It is this, Taylor contends, which suggests a link with Tarski's Convention (T).\textsuperscript{14} What link?

'There a theory of sense for $L$ so chooses its $\xi$ that its version of (I) has the form (II) $s$ is true-in-$L$ $\leftrightarrow$ $p$, then it is construable also as a theory of truth for $L$ recursively assigning truth-conditions to each of $L$'s indicative sentences. Construed as a theory of sense, it gets its content assignments right just in case (construed as a theory of truth) it is adequate by the lights of Convention (T).'\textsuperscript{15}

'Adequacy by the lights of Convention (T)' is no standard which Julio can avail himself of to select between equally true truth theories for Diana's language, since, in Davidson's reversal of Tarski's procedure, it is truth which is taken as primitive and used to illuminate translation. Cleaving to extensionality, we were forced to acknowledge that a theory of sense of the form (II) $s$ is true-in-$L$ $\leftrightarrow$ $p$ admitted as values for metalinguistic $p$ sentences which did not translate $s$. So what justifies the construal of instances of (II) as 'meaning-giving' clauses at all?

\textsuperscript{13} Where $s$ is non-indicative, we take the content of $s$ as that which the theory assigns to its indicative transform.

\textsuperscript{14} When our $\xi$ is a Tarskian truth-predicate 'is true-in-$L$ iff', $\xi$ will for each sentence $s$ of $L$, provide a canonical proof procedure such that:

(i) At any stage at which the semantic predicate(s) advertised to in $\xi$ is (are) applied to a complex sentence of $L$, we use the axiom for the last constant used in the construction of the sentence to determine the result of applying the predicate(s) to less complex sentences of $L$.

(ii) At any stage at which the semantic predicate(s) is (are) applied to a simple $L$-sentence, we use that sentence's axiom to eliminate that use of the predicate.

(iii) At any stage at which the semantic predicate(s) does (do) not appear, we stop.

\textsuperscript{15} Modes p.s.
Certain answers to this question which seem to enjoy some currency seem to me to be wrong. Thus McDowell at 'Truth-conditions. Bivalence and Verificationism' appeals to the 'fact' that appending 'is true' to a designation of a sentence is equivalent to using the sentence. But as we saw in §1.2, this alleged equivalence, when construed as one of synonymy, is not at all obvious, and when the connective between $s$ and $p$ is an extensional one as in (II), we precisely cannot construe an assertion that $s$ is true as an assertion that $p$.

Platts follows McDowell in this:
'The effect of the disquotation device, the truth predicate, is to produce a sentence which can be used to say the very same thing, to perform the very same propositional acts, as could the original sentence's prior to quotation or designation.'

This cannot be right. Suppose Diana knows from a Spanish-speaking friend that 'La terra si muove' expresses a truth but does not know which truth. Julio ardently proclaims 'La terra si muove' as he clasps her hand. Only if she can interpret his utterance can she use the disquotational property of the truth predicate to construe his utterance as, in primary redescription, an assertion that the Earth moves. We may be able to appreciate that the truth-predicate functions as a disquotation device when we have an adequate theory of sense, but the problem precisely is to derive such a theory from a theory of truth alone.

Taylor's attractive claim above that a truth-theory satisfying $\text{Con}(T)$ will eo ipso be construable as a theory of sense thus awaits a better defence than the ones considered so far.

---

§1.3.2 Indexicality and truth theory.

Indexical utterances will provide crucial evidence about the meanings of a speaker's words which no plausible theory of meaning for that speaker's language can afford to overlook. Yet indexicality (and ambiguity) present difficulties for Davidson's project.

Tarski defined truth for an interpreted formal language. A truth definition for a natural language requires its sentences to be paraphrased into some suitable canonical idiom. Assuming such paraphrase to be possible, object language sentences will not be translated by natural language metalinguistic counterparts in the truth theorems for those sentences, but instead by wffs in the formal language (normally a natural extension of first order predicate calculus). Each OL sentence will be paired with a semantic representation in the canonical idiom which gives its 'logical form'. Davidson's assumption is in Barry Taylor's words that:

'... underlying a natural language like English, there exists a formal language ('Base English') in which every surface sentence of English finds at least one paraphrase (a 'base paraphrase'); surface sentences are further supposed linked to their base paraphrases by a series of meaning-preserving Chomskyan transformations, and the task of constructing a theory of meaning for English identified with that of finding a recursive truth-theory for Base English meeting appropriate constraints.'¹⁸

The problem caused by indexicality and ambiguity for such a project is that on Tarski's conception, truth is a predicate of sentence-types, yet the truth of a sentence with indexical or ambiguous constructions depends on the context in which it is uttered. So the truth-predicate cannot be straightforwardly applied to those sorts of sentences.

¹⁸ 'Truth theory for indexical languages' p.182.
As Taylor notes\(^{19}\), there are two broad responses one can make to this difficulty—either:

(a) still regard truth as a unary predicate but no longer of sentences; or

(b) still think of truth as a predicate of sentences but as relativised to appropriate parameters.

I wish to outline some proposals for dealing with the problem of indexicality which adhere to either (a) or (b) above and to briefly sketch two approaches I find promising.

The leading idea of our first proposal is a straightforward one: the truth-value of an indexical sentence \(s\) is alleged to depend just on those objects which its indexical terms refer to in a given context.\(^{20}\) I shall call this the 'points of reference' approach, following Scott and Taylor\(^{21}\), and label it \((P_1)\).

Then, according to the proponents of \((P_1)\), truth is to be predicated of type-sentences relativised either to 'indices' \(^{22}\) (octuples consisting of worlds, times, places, speakers, audiences, sequences of indicated or demonstrated objects, discourse segments, and sequences of free variables) or more generally 'contexts' \(^{23}\) which leave it open which features of context are fixing the parameters in any given instance. Then the intuitive suggestion is that:

\(^{19}\) loc.cit.

\(^{20}\) More precisely, the features of context on which the truth-value of \(s\) depends are to comprise a finite sequence of objects whose \(i^{th}\) element in the given context is the entity the \(i^{th}\) occurrence of the indexical term \(e\) in \(s\) denotes.


\(^{22}\) As in Scott loc. cit. and Montague, R. 'Pragmatics and Intensional Logic' in his *Formal Philosophy*, Thomason, R (ed.) New Haven, Conn.: Yale University Press, 1974.

(1) 'I am happy today' is true at C(p,t) iff p is happy in C at t.
where C is a context or an index containing at least the speaker p and time of
utterance t.

But there is a difficulty for this approach which William Lycan notes.
Consider the sentence 'I am silent now'. Then, by the lights of (P₁),

(2) 'I am silent now' is true at C(p,t) iff p is silent in C at t.
While it accommodates the fact that 'I am silent now' is true of a p who is not
now speaking, (2) does not even predict, let alone explain, why that sentence
is false whenever uttered (and thus would never be volunteered seriously
by a p who understood English). Moreover, since any utterance of this
sentence is necessarily false on the lips of its utterer, it would seem that,
contra Davidson, truth cannot be a property of utterances.

Tyler Burge writes:
'sentences containing a demonstrative element are on some occasions not just true or false of
objects but are true or false. Thus we must alter the traditional concept that truth is a property
only of closed sentences.'

(P₁) ignores this fact - indexical sentences are held to be true at some index
or context, false at others, but the property of falsity tout court which 'I
am now silent' acquires whenever I utter it receives no explanation.

The second approach is the one Davidson favours. I will refer to this as the
'speaker demonstration' approach and label it (P₂). (P₂) seeks to explain how
the elements of an index or point of reference are determined as those

24 Lycan Logical Form p.51.
25 'Demonstrative constructions, Reference and Truth' p.213 Journal of Philosophy vol.71,
1974.
26 Lycan loc. cit. pp.51-52
entities which a speaker demonstrates at a time. To take the simplest example:

(3) "This is hot'. p, t is true iff the object demonstrated by p at t is hot.

The idea is that truth is a predicate of ordered n-tuples consisting of a sentence-type, a speaker, a time, that sequence of entities demonstrated by the speaker at that time.

Taylor notes a problem for (P₂) caused by cases of demonstrative equivocation wherein a speaker demonstrates two or more distinct entities by means of a single utterance.²⁷ In his example, I am waylaid on the telephone by a long-winded colleague and my wife slips me a note asking if I want her to make noises as of the arrival of a guest. With a wink in her direction I say 'That's a good suggestion', ostensibly commending my colleague's trite proposal, and at the same time sincerely welcoming my wife's canny one.

As Taylor acknowledges, (P₂) can handle this by including an audience parameter in the n-tuple of which truth is predicated - thus:

(4) "That's a good suggestion'. p, t, A is true iff p judges A's suggestion to p at t to be good.

So the sentence 'That's a good suggestion' is true on my lips when directed to my wife (since the object demonstrated by me to her was one I judged to be a good suggestion) but false on my lips when directed to my colleague's (since the object I demonstrated to him was one I judged not to be a good suggestion).

Taylor thinks that (P₂) will not be able, in the face of demonstrative equivocation involving a single audience, to find a:

... set of features whose addition to the familiar parameters of utterer and time will suffice in the necessary way to disambiguate all cases of demonstrative equivocation.²⁸

²⁷ loc. cit. p. 185.
²⁸ loc. cit. p. 186.
This is surely right. In the case of a single audience, we will have to evaluate an equivocal demonstrative utterance at its point of reference without having any systematic means of determining the elements of its point of reference.

Thus if my wife shows me a photograph in the newspaper of a chemical factory belching pollutants into the atmosphere and I, with one scornful eye on the picture and another admiring eye on her in her new dress, say 'That's a lovely sight', \((P_2)\) will have to posit another member of the n-tuple in an unsystematic fashion:

\[(5) \text{"That's a lovely sight", } p, A, o, t > \text{ is true iff the object } o \text{ demonstrated by } p \text{ to } A \text{ at } t \text{ is adjudged by } p \text{ a lovely sight.}\]

Whence 'That's a lovely sight' said of \(o=\) my wife in her new dress is true and of \(o=\) the picture of the factory, is false.

So it seems that \((P_2)\) could never wholly displace \((P_1)\) - we can never reduce all features of a context on which the truth-value of an indexical sentence might conceivably depend to a finite set of parameters circumscribable in advance.

We should note in passing that \((P_2)\) can account for 'I am now silent'. If an utterance of the sentence-type 'I am now silent' were to be made by \(p\) at \(t\) then it would be true if and only if \(p\) is silent at \(t\). So the clause:

\[(6) \text{"I am now silent", } p, t > \text{ is true iff } p \text{ is silent at } t\]

together with the standard clause for negation will give us the correct result.

Weinstein's and Burge's method of conditional assignments looks more promising. I will refer to this as \((P_3)\). The intuitive idea behind this proposal is this:

If \(u\) is a speaker and \(\sigma\) a sequence of objects demonstrated by \(u\) in context \(C\) by the indexical expressions contained in \(s\) in the order in which they occur then \(u\)'s utterance of \(s\) is true iff \(\sigma\) satisfies the result of replacing the indexical expressions in \(s\) by corresponding free variables
Our T-sentences for indexical sentences thus become conditional in form. Ever since Wittgenstein, the Land of Bare Particulars has had few philosophical visitors, and rightly so - Wittgenstein taught us that there can be no demonstration of a bare something, that we must demonstratively refer to items under some aspect or other. As a result of this, it seems a mistake to represent 'That F is G' as (That x)(Fx & Gx) in canonical first order idiom. In order to account for the asymmetry\(^{29}\) induced by the feature under which we subsume a demonstrated object, it is therefore desirable to represent the demonstrative 'that' quite generally as a binary quantifier so that 'That F is G' is represented as '(that x)(Fx;Gx)'.

In order to discuss (P\(_3\)) adequately, we will need to introduce some terminology. For simplicity's sake, I will only consider the semantic effect our binary demonstrative quantifier has upon a first order language \(L\).\(^{30}\)

(i) Let the *indexical extension* \(L_{I}\) of \(L\) be the language obtained from \(L\) by adding to it the quantifier '(that x)(Fx;Gx)'.

(ii) Let \(\sigma, \sigma', \ldots\) be metalinguistic variables ranging over finite sequences of objects, where the \(i^{th}\) element in any such sequence is the item which in the context \(C\) satisfies the \(i^{th}\) occurrence of the binary demonstrative quantifier '(that x)(Fx;Gx)'.

(iii) Define the *indexical degree* of a wff \(A\) as the number \(n\) of occurrences of the demonstrative quantifier '(that x)(Fx;Gx)' (where \(n=0\), we say that \(A\) is non-indexical).

(iv) Call a finite sequence \(\sigma\) *apt for* a wff \(A\) just in case the length of \(\sigma\) is at least as great as \(A\)'s indexical degree.\(^{31}\)

Now we wish to account for a complex notion of speaker-demonstration - to wit, u's demonstrating \(x\) *as an F*. Taylor suggests as a rough first approximation for this notion:

---

\(^{29}\) That is, whether we say 'That spy is a fool' or 'That fool is a spy'.

\(^{30}\) Here, with slight alterations in symbolism, my terminology and procedure follows Taylor's.

\(^{31}\) Cf Taylor op cit.
"That F is G" is true relative to u, t, and non-empty \( \sigma \) iff at his first demonstrative utterance during t, u demonstrated \( \sigma_1 \) as an F; and moreover, \( \sigma_1 \) is G. 32

But this, Taylor notes, is unsatisfactory because F, though occurring extensionally in "That F is G", occurs intensionally in the paraphrase. Taylor therefore replaces the intensional version with an extensional analogue "u demonstrates x as a member of \( \alpha \)." 33

So let us symbolise "x is an item demonstrated by u as a member of \( \alpha \) in his ith utterance during t of the demonstrative quantifier "(that x) (Fx; Gx)" as 'Dem$_i$(u, t, x, \( \alpha \)).' Then we can lay down an adequacy condition for a truth theory for LI cast in a metalanguage ML containing L together with Tarski's auxiliary resources as follows:

32 loc cit p. 192.
33 Where u demonstrates x as an \( \alpha \) if for some F, \( \alpha \) is the set of F-things and u demonstrates x as an F.
An adequate truth theory for LI in ML will be one yielding as a theorem for each closed wff \( A \) of indexical degree \( n \) an instance of the schema:

\[ \sigma \text{ is apt for } \Omega \rightarrow. \]

\[ T_{LI}(\Omega, u, t, \sigma) \leftrightarrow \forall v_1 \ldots \forall v_n (v_1 = \sigma_1 \land \ldots \land v_n = \sigma_n \land \Phi) \]

where (a) The schematic letter '\( \Omega \)' is replaced in both of its occurrences by a structural description in ML of \( A \).

(b) The schematic letter '\( \Phi \)' is replaced by the last member \( B_m \) of a shortest sequence \( B_1, B_2, \ldots, B_m \) such that:

(i) \( B_1 \) is the demonstrative-standard equivalent \(^{34} \) of \( A \);

(ii) \( B_m \) is non-indexical;

and (iii) for each \( i \) \((1 \leq i \leq m)\), \( B_{i+1} \) results from \( B_i \) by replacing some part of the form 'that \( v_k(C;D) \)' (where \( C \) and \( D \) are non-indexical) by \( 'Dem_k(u,t,v_k,C) \land D' \). \( \ldots (AC)^{35} \)

Taylor sees a problem with this conditional assignment approach - he thinks that it will not be able to adequately account for sentences involving quantification into descriptions embedded inside the binary demonstrative quantifier 'that x)(Fx;Gx)'.\(^36\) For example, when \( A \) is the wff:

\( (\exists y)(Loathes(y, (that x) Denigrates(x,y)) \)

which, Taylor claims, is one formalization of 'There is someone who loathes that denigrator of his', any truth theory with the consequence:

\[ Dem_1(u,t,\sigma, (x| Denigrates(x,y))) \rightarrow. T_{LI}(A,u,t,\sigma) \leftrightarrow (\exists y)(Loathes(y,v_1)) \]

will be inadequate because the variable '\( y \)' in the set-abstract

---

\(^{34}\) Where \( A \) is a wff of indexical degree \( n \), its demonstrative-standard equivalent is the wff \( A^* \) formed from \( A \) by:

(i) replacing every variable \( v_i \) at each of its occurrences in \( A \) unbound by a demonstrative quantifier by \( v_{i+n} \); and

(ii) where \( v_{jk} \) is the variable associated with the \( k^{th} \) occurrence of the demonstrative quantifier in \( A \), replacing \( v_{jk} \) by \( v_k \) at each of its occurrences bound in \( A \) by that occurrence of the quantifier.

\(^{35}\) This is Taylor's adequacy-condition for truth-theories for indexical languages, [Tiv], at p.193 loc. cit.

\(^{36}\) loc. cit. p.195.
'(x) Denigrates(x,y))' is not bound by the existential quantifier in the consequent of the conditional.

I wonder if we may not be able to bypass this problem. Insofar as this sentence is semantically coherent at all, it seems odd that we should regard the predicate 'is a denigrator of y' as used to accomplish an identification of x when we do not know who the y is. Hence there appears to be a ground for not thinking of the situation as one in which x is demonstratively identified as a denigrator of some y but rather as one in which his being a denigrator (of some y) is part of the information which u provides about the demonstratively identified x in question.

Then how is x demonstratively identified? Presumably, as either a man or a woman. So we then reparse the problematic sentence as:

'There is someone who loathes that person who is a denigrator of his.'

Then we will have:

Dem1(u,t,u,(x) Person(x))). \rightarrow T\text{LI}(A,u,t,u) \rightarrow (\forall y)(\text{Denigrates}(\sigma_1,y) \& \text{Loathes}(y,\sigma_1)) ... (7)

and the adequacy condition (AC) on truth theories for LI in ML will be capable of being satisfied.

How will (P) fare with 'I am now silent'? We would need to enrich LI to include token-reflexives to make sentences such as this one amenable to a precise specification of truth-conditions. I shall not attempt to provide such a specification here. Instead, I will simply sketch in rough outline how those truth-conditions might look:

Suppose we represent the sentence 's is a sequence of denotata of the token-reflexives \(\tau_1, \tau_2, ... , \tau_n\) in the order in which they occur in u's i-th utterance of A during t' as 'Tck_i(A,u,t,s)'. By an obvious intuitive extension of the
strategy with demonstrative expressions, the conditional assignment
proponent will want to articulate the truth-conditions for 'I am now silent'
along the following lines:

\[ \text{To}_{t_1}(\text{I am now silent}, u, t, <u, t>). \]

\[ \text{T}_{t_1}(\text{I am now silent}, u, t, v)(\exists v_1[v_1 = u \& v_2 = t \& v_1 \text{ is silent at } v_2]) \ldots (8) \]

If we now add to (8) the axiom:

\[ (VA)(Vu)(Vt)(\text{if } u \text{ utters } A \text{ at } t \text{ then } u \text{ is not silent at } t) \quad (Ax) \]

we can derive the conclusion that if \( u \) utters 'I am now silent' at any time
then his utterance is false. Moreover, just because indexical sentences are
regarded as open sentences when they are not being uttered in a specified
context, (P3) can explain why 'I am now silent' is true of any person who is
not speaking at the relevant time.

So, it appears as if (P3) can adequately explain the semantic features of this
sentence along with other sentences containing demonstrative and indexical
terms.

Lycan has suggested a far simpler method of treating indexicality. He
invokes an assignment function \( V \) that takes the indexical sentence and its
respective context as input and yields as output a sequence of denotata
for the sentence's indexical elements in that context.\(^{37}\) In fact, \( V \) is to be
thought of as a complete valuation function for all the referring terms of the
OL, including non-indexical individual constants if any. The idea is to prise
pure semantics away from complex psycho-social factors. Then a 'common
sense' pragmatic account tells us how to compute \( V \) for personal pronouns, a
theory of ostension gives us a recipe for computing \( V \) for demonstratives
such as 'This', 'That', 'Here' etc. Thence:

\(^{37}\) Logical Form pp.54-55.
(9) 'I am now silent' is true in C iff $V('I', C)$ is silent at $V('now', C)$

The pragmatic theory will explain why 'I am now silent' is false whenever uttered because it will make manifest that for any C, $V('I', C)$ is speaking at $V('now', C)$.
§1.4 RADICAL INTERPRETATION

Not every truth theory for L which meets the adequacy constraints on theories of meaning can be a theory of sense. How then can we ever justify construing certain T-sentences as 'meaning-giving' ones? We saw in §1.3 that certain proposed justifications for such a construal were defective. McDowell believes that 'deviant' truth theories for L will be eliminated when, in the context of radical interpretation, we subject our attributions of content to N's utterances and suspected psychological states to the regulative ideal of making the best rational sense of N's actions.1 Perhaps. But we should recall that Quine, the father of the enterprise of radical translation, drew just the opposite conclusion.2

Quine's 'radical translator' revolutionised the theory of meaning. Prior to Quine, meanings had been assumed to be both determinate and indispensable in the theory of meaning. A translation between two languages was deemed adequate just so long as synonymy between designated pairs of expressions of those languages was preserved. Quine challenged these dogmas and inverted conceptual priorities in an effort to show that: 'the notion of linguistic meaning is not a scientifically useful one.' 3

Quine thought that any illumination of the notion of linguistic meaning must come through eliciting empirical constraints on the translation into one's own

---

2 Namely, that translation is indeterminate. McDowell discusses indeterminacy of translation in 'Truth Conditions, Bivalence and Verificationism' pp.65-66 but admits that he has no answer to Quine.
tongue of a completely alien language. His field linguist started by correlating certain of N’s utterances (occasion sentences) with the environmental stimuli prompting N’s assent to those sentences when queried (stimulus meanings). Translation then commenced at the point where the linguist found sentences in his own language with the same or almost the same stimulus meanings as ones in the jungle language.

Quine’s story and its conclusion are by now familiar – only a minute number of pairs of sentences of the jungle language and the linguist’s own will be stimulus synonymous (prompted by the same stimuli); the translation of truth-functional sentential connectives will be fairly determinate, the translation of the vast majority of sentences, indeterminate. It is an illusion to think otherwise ... or so Quine says and Davidson agrees.4

Now we have agreed with Davidson that Julio will be able to acquire Charles’ ability to communicate with Diana if he knows how to deploy the outputs of a theory of sense in concert with those of a theory of force to produce primary redescriptions of her utterances. We know that a truth theory for Diana’s language which passes Con(T) (suitably modified as suggested in §1.3 to account for indexicality) will in some sense ‘translate’ or ‘give the content of’ her utterances. But how are we to obtain such a theory? If our aim in looking at the plight of the radical interpreter was to cast light on meaning, then we had better not, when we come to pair structurally designated jungle language sentences with metalinguistic sentences, just assume that we can

spot correct translation when we come across it, since translation supposedly preserves meaning.

Davidson does not assume this. He begins by surrendering the requirement that the metalinguistic sentence on the RHS of the biconditional in the T-sentence translate the designated object-language sentence on the LHS. Object language sentences are then paired with co-extensive metalinguistic sentences - i.e., Davidson simply requires that T-sentences be true. Then, by laying down enough formal and empirical constraints on the resultant truth theory, he seeks to elicit for each OL sentence s what is in fact the translation of s on the RHS of its corresponding T-sentence. Thus he writes:

> 'The present idea is that what Tarski assumed outright for each T-sentence can be indirectly elicited by a holistic constraint.'

Davidson's hope is that a holistic constraint on the theory can do what piecemeal modifications to the T-sentences themselves (in particular, to their constituent biconditionals), failed to do - viz. produce empirically testable translations of the sentences of the object language. What evidence could we use to construct such a theory?

Davidson argues that N's fine-grained psychological attitudes cannot provide the evidence a radical interpreter R needs to test his truth theory for N's language. For, he claims, we cannot have access to the contents of N's attitudes in advance of constructing a theory of meaning for his language. I will call this (Quinean) thesis the Inseparability of Belief and Meaning Thesis (IBMT). Because of (IBMT), Davidson argues that we must look elsewhere than N's psychological attitudes for our evidential base. Davidson believes that certain of N's sentential attitudes can provide the required

---

5 'Radical Interpretation' loc. cit. p. 139.

6 Compare his remark 'Interpreting a speaker's intentions, beliefs and words are part of a single project.' p. 127 in 'Radical Interpretation'; cf 'Belief and the Basis of Meaning' p. 144 on ascribing a belief in the existence of a largest prime number.
evidence. These attitudes are N's attitudes to the truth of his own sentences. Without knowing what s means, we can know that N holds s true or prefers it to be true or intends that it be true etc. Davidson contends. The thought apparently is that because holding-true is not a propositional attitude, it does not fall foul of (IBMT).

There is surely a problem, however, in interpreting a form of behaviour as full assent to the truth of a sentence. Davidson just assumes that there is a recognisable form of behaviour which expresses full and unconditional assent, but it is possible that N has several different kinds of assent, in particular he may think that a sentence is merely assertible, in fact he might only have the attitude of holding certain sentences assertible. So it is not the content of the given sentence but N's intensional attitude to it which is problematic here since Davidson cannot just take 'true' in 'holding-true' to be semantically inert.
If R is not sure that Karl holds 'Es regnet' true when (and only when) it is raining, he will have a hard time establishing the truth of the T-sentence for Karl's language:

"Es regnet" is true as uttered by Karl at t iff it is raining near Karl at t.

(A)

For, in order to then verify (A), R would have to know both what 'Es regnet' means and whether it really was raining near Karl at t. Provided R can know that it is truth which N imputes to the sentences he (N) assents to, R can avoid this worry. On the LHS of the T-sentences, R need only accept N's opinion, on the RHS, his own. So if at time t and location x R believes that it is raining and Karl assents to 'Es regnet', R can take this as (prima facie) confirmation of (A) and thus of one consequence of his truth theory for Karl's language. But can R know that Karl holds 'Es regnet' true rather than, say, assertible?

I think Davidson can answer this question in the affirmative if he takes, as I suggested in §1.2 he ought to take, truth to be primitive in accounting for the cognitive modalities of propositional attitudes. Then, Karl's judging that it is raining is to be explained as a judgement that the thought that it is raining is true - judging is judging-true (where truth again is not inert but primitive). But judging is to be explained as judging a certain content and not a sentence true. So the price of this defence for Davidson is the concession that holding-true is, after all, subject to (IBMT).

7 Which would of course reduce Davidson's program to absurdity - since to even test the truth theory, we would first have to know the meanings of all OL sentences (together with the truth of every one of the metalinguistic sentences which we proffer as giving their truth-conditions).

8 The claim would be that Karl's assertion of 'it is raining' or his assent to the truth of 'it is raining' can only be construed as instances of Karl's holding of 'Es regnet' that it is true because Karl uses this sentence to express the judgement that it is raining.
There is another reason why we might doubt whether T-sentences are really testable (and thus whether a truth theory is genuinely empirical) - this is that the metalanguage in which the truth predicate is being recursively characterised will not be a natural language but some appropriately regimented (quasi) formal language. Thus, if Davidson's analysis of action sentences is correct, a truth theory for English might contain clauses such as:

"J^o^n^e^s\ b^u^t^e^r^e^d\ t^h^e\ t^o^a^s^t\ l^n\ t^h^e\ b^a^t^h^o^m^"\ is\ true-in-Base-English\ ↔\ (\exists e)\ [B(J,e)\ &\ t(b,e)]^\ ...\ (B)

Following Lycan, let us call such clauses 'impure' T-sentences.9 By way of contrast, a 'pure' T-sentence for the target sentence will be:

"J^o^n^e^s\ b^u^t^e^r^e^d\ t^h^e\ t^o^a^s^t\ l^n\ t^h^e\ b^a^t^h^o^m^"\ is\ true-in-English\ ↔\ Jones\ buttered\ the\ toast\ in\ the\ bathroom\ ...\ (B*)

Clearly, only the pure T-sentences are directly confirmable by English speakers. But Base-English sentences can be mapped onto surface English sentences by a sequence of meaning-preserving syntactic transformations - it is a simple matter to then prove that: s is true-in-English ↔ s is true-in-Base-English. We can therefore regard empirical confirmation of pure T-sentences such as (B*) as providing evidence for the truth of impure T-sentences such as (B) only on the assumption that the syntactic theory mapping logical forms onto surface structures is correct. Alternatively and holistically, the evidence confirms a truth theory for Base-English together with a (Chomskyan) syntactic theory.10

---

9 Logical Form p.31.
10 I think this is the thrust of Gilbert Harman's position which Lycan discusses and rejects at pp.299-300 of Logical Form. Lycan seems to require that we somehow be able to test the truth-theory by itself, but surely any plausible theory of meaning must be developed in tandem with a syntactic theory.
It seems to me then that a truth theory for \( L \) will be confirmable in the light of the semantic intuitions of \( L \)-speakers - where the truth theory predicts an implication between \( s_j \) and \( s_k \), \( L \)-speakers should disallow the possibility that the former sentence might be true without the latter being true also\(^{11}\); where the truth theory assigns two (or more) readings to 'Information about the papal mass can be found on page three' in Base-English, English speakers should detect an ambiguity in that sentence and so forth.

Davidson stresses the fact that \( T \)-sentences represent empirical generalisations\(^{12}\) - we can test a \( T \)-sentence as we test any law: by noting that there is an accretion of positive confirmation of the law from instance to instance, by seeing that it supports counterfactuals etc. Thus, 

"snow is white" is true \( \leftrightarrow \) grass is green' \( \ldots (C) \)

is no law as julio can quickly discover if he can but convey to English-speaking Charles or Diana that he wishes to know whether 'snow is white' would still be true were grass pink.\(^{13}\)

\(^{11}\) Provided the implication is sufficiently obvious to them of course!

\(^{12}\) Cf. Introduction to Truth and Interpretation pp.xiv,xviii; compare also Davidson's claim at 'Reply to Foster' p.174: 'A theory that passes the empirical tests is one that in fact can be projected to unobserved and counterfactual cases, and this is apparent to anyone who knows what the evidence is and how it is used to support the theory.' It is clear that the theory's \( T \)-sentences for indexical sentences represent empirical generalisations - since \( R \) can discover that an empirical correlation holds between Karl's holding 'Es regnet' true and its raining. However, for eternal sentences there is no way of gaining evidence for any such correlation since we only have the set of eternal sentences held true by Karl together with the facts as \( R \) perceives them, and no one correlation is preferable to any other. \( R \) can make no progress on the meanings of Karl's eternal sentences without invoking the whole truth-theory.

\(^{13}\) This is not such a difficult project. For example, a drawing of white snow on green grass followed by one of white snow on pink grass with the query 'snow is white?' accompanying each diagram might conceivably achieve this. But we should note that once one says that the experiment provides evidence for a particular \( T \)-sentence, one is already interpreting the biconditional as a non-extensional one. If the biconditional is extensional, the experiment is irrelevant. Thus Davidson comments: 'So a theorem like "Schnee ist weiss" is true in the mouth of a German speaker if and only if snow is white' has to be taken not merely as true, but as capable of supporting counterfactual claims. ... How much of a concession this is to intensionality depends, I suppose, on one's analysis of the concept of law.' 'Introduction' p.xiv in Truth and Interpretation.
For sentences R judges to contain observational predicates (with featureplacing expressions or mass terms for example), ones for which perception or ostension seem relevant, R would do well to start the process of interpreting N's speech by attributing the same perceptual beliefs as he himself holds, taking certain context-sensitive utterances as expressing those beliefs. But unless we simultaneously attribute truth-conditions to the corresponding sentences such attributions will offend (IBMT), so it is time to look at that constraint. Davidson writes:

'radical interpretation cannot hope to take as evidence for the meaning of a sentence an account of the complex and delicately discriminated intentions with which the sentence is typically uttered ... the reason is that interpreting an agent's intentions, his beliefs and his words are part of a single project, no part of which can be assumed to be complete before the rest is. If this is right, we cannot make the full panoply of intentions and beliefs the evidential base for a theory of radical interpretation.'

Davidson seems to me to be attacking a straw man in this passage. For why must the evidential base of psychological attitudes be assumed 'complete' before we attribute meanings to N's words? Which Gricean ever held that we need the 'full panoply' of intentions and beliefs before venturing guesses as to what N means by his utterances? As Jonathan Bennett has argued, the Gricean can, just as readily as any Davidsonian, avail himself of a 'Bootstrap Strategy' - we hazard some guesses as to what N's beliefs, desires and

14 Thus, whether 'Gavagai Yamoos!' means 'the rabbit has disappeared' or some unrepealable expletive in Quindi pursuant upon a rabbit disappearing before Quindian N could steady his aim, R will be justified in ascribing to N the belief that a rabbit has escaped N's arrow, as well as an intention to kill it, a desire to feast upon rabbit meat etc. With this, Davidson agrees.

15 How do we know that N at 22n above has the concept rabbit rather than small edible furry thing when we impute to him the belief that a rabbit has just escaped his arrow? The only way we can decide this according to Davidson is to solve for the meanings of his words as well as his beliefs and attitudes.

16 'Radical Interpretation' loc. cit. p. 127. I am puzzled by the phrase 'an account of (the complex and delicately discriminated intentions).' ... How can an account of psychological states be evidence for any empirical theory of interpretation. Isn't it the states themselves which provide the evidence?

intentions might be on the basis of N's behaviour and the prevailing environmental conditions. Such initial attributions of attitudes are very tentative and rough but this is no problem - we proffer them anyway even though we know them to be only approximately correct, if correct at all. This gives us a basis for rationalising N's actions and for construing the mode if not the fine content of some of his utterances and we use this subsequent information to discriminate the initial coarse-grained attributions more finely.

Moreover, there seems little in Davidson's holistic strictures on content attribution in (IBMT) to worry a Gricean. The Gricean takes speaker's meanings to be special sorts of intentions - 'meaning intentions' or 'M-intentions'. That a speaker's meanings (M-intentions) are inextricable from the rest of his psychological attitudes is, for the Gricean, an obvious result of the holism of the mental which prevents the attribution of any single type of mental state outside of a battery of supporting psychological attitudes to the agent. Hence, metaphysical conclusions about the nature of meaning and belief simply do not follow from (IBMT) alone, since that thesis can be construed as little more than a platitude about psychological attribution - in particular, (IBMT) does not itself show that meaning is conceptually irreducible to intention (even if there can be no epistemic prising apart of meanings and attitudes within the context of radical interpretation).

18 The clearest exposition of such a strategy from a broadly Gricean point of view is undoubtedly David Lewis' 'Radical Interpretation' Synthese 23, 1974, pp.331-344.

19 Compare Michael Dummett's appraisal at 'What is a Theory of Meaning?' p.127 in Guttenplan, S (ed.) Mind and Language, Oxford University Press, Oxford, 1975, pp.97-138: 'Holism in respect of how one might, starting from scratch, arrive at a theory of meaning for a language ... is, as far as I can see, unobjectionable and almost banal. It is certain that Davidson intends his holism as a doctrine with more bite than this.'
Still, I believe that Davidson is right to hold against Grice that semantics is not reducible to psychology. My reason for thinking this is that, as I argued in §1.2, we require an irreducible semantic notion - viz. truth - to even start to account for the nature of the various cognitive modalities under which contents are entertained.

I have agreed with Davidson that a truth theory is an empirically testable theory, but there is at least one difficult problem with Davidson's methodology:

Davidson suggests that R have available to him all the sentences N holds true or prefers true. The determinants of Karl's holding 'Es regnet' true are:

(a) What 'Es regnet' means; and (b) What Karl believes.

Davidson claims that as a result of (IBMT), we cannot have access to either without the other. He compares his problem with the problem of solving for two variables given only one equation, e.g. \( x^2 - y = 9 \).

But there is an asymmetry between R's variables in (a) and (b) not reflected in the equation in \( x \) and \( y \) above: given a knowledge of the meaning of 'Es regnet' together with sufficiently many instances of holding 'Es regnet' true, R can work out what Karl believes on the appropriate occasion, but the converse does not hold.
For even if R knows that Karl believes that it is raining on the occasions when it is, he still is in no position to determine the meaning of 'Es regnet', since according to Davidson, R also has sufficiently many instances of N’s holding (vast numbers of) other sentences true on just those occasions when it rains. R can therefore only match the requisite belief with 'Es regnet' if he knows what, ex hypothesi, he does not know - viz. that 'Es regnet' expresses the belief that it is raining.

Indexical utterances undoubtedly supply a rich source of data for R and because the presence of demonstrative and token-reflexive expressions in a natural language is all-pervasive, their contribution to the testing of the truth theory is invaluable. In recognition of this, Davidson writes:

'One important, indeed essential, factor in making a truth theory a credible theory of interpretation is relativisation to speaker and time ... a theory that makes the right sentences true at the right times and for the right speakers will be much closer to a theory that interprets the sentences correctly than one that can ignore the extra parameters.'

But, a large body of sentences of a natural language do not contain indexical expressions. So how is R to gather evidence as to what these sentences mean? While it is difficult to see how to complete "'It is raining" is true as uttered at t by N ↔ ... " with anything other than 'It is raining near N at t'[^21], it is simple to complete "'It is always raining somewhere at some time" is true ↔ ... " which makes no reference to particular spatio-temporal regions with metalinguistic sentences which have nothing to do with the weather (e.g. 'Hitler invaded Poland').

[^20]: 'In Defence of Convention T' pp. 74-75 in Truth and Interpretation, pp.65-75. I think we can say something much stronger than this: viz. a theory that 'ignores the extra parameters' is an impossibility!
[^21]: Difficult, but not impossible as Foster and Loer proved.
Davidson is not explicit on how we might advance from occasion sentences to the rest.\textsuperscript{22} I wish to discuss two features of his method which are crucial in effecting such a transition. The first is his holism; the second, is his version of the Principle of Charity, or, as he prefers to now call it, the Principle of Rational Accommodation (PRA). First, holism.

\begin{quote}
\textsuperscript{22} Witness his remark at 'Radical Interpretation' p. 136: 'The process of devising a truth theory for an unknown native tongue might in crude outline go as follows. First, we look for the best way to fit our logic, to the extent required to get a theory satisfying Convention T, on to the new language; this may mean reading the logical structure of first-order quantification theory (plus identity) into the language, not taking the logical constants one by one, but treating this much of logic as a grid to be fitted on to the language in one fell swoop. The evidence here is classes of sentences always held true or always held false by almost everyone almost all of the time (potential logical truths) and patterns of inference. The first step identifies predicates, singular terms, quantifiers, connectives and identity; in theory it settles matters of logical form. The second step concentrates on sentences with indexicalis, those sentences sometimes held true and sometimes false according to discoverable changes in the world. This step in conjunction with the first limits the possibilities for interpreting individual predicates. The last step deals with the remaining sentences, those on which there is no uniform agreement, or whose held truth value does not depend systematically on changes in the environment.'
\end{quote}
Suppose R is unsure whether Karl means 'It is raining' by 'Es regnet' or 'It seems to me that it is raining'. Attention to the way 'Es regnet' embeds in compound sentences and to the semantic effect various sentential operators have upon the sentence should enable R to determine that Karl means that it is raining by 'Es regnet'. For all he need do is note the effect Karl's negation operator has upon 'Es regnet' - Karl will assert 'Es regnet nicht' only when he has evidence that it is not raining and not when he simply has no evidence whether it is or not. Quite generally, the holistic enterprise of providing a truth definition for Karl's whole language will impose severe constraints on the acceptability of putative translations of Karl's sentences - by attending to the semantic effect modal, tensed, counterfactual and propositional attitude operators (as well as the more familiar truth-functional operators) have upon various candidates for the translation of a given sentence s, R will be able to eliminate many erstwhile plausible competitors to the conventionally accepted one. Moreover, the fact that the derivation of T-sentences is required to be canonical will serve to rule out certain undesirable T-sentences which, though true, are not derivable.

Insofar as Davidson's theory accommodates the notion of meaning at all, the meaning of a sentence is not determined just by its truth-conditions but by the system of inferential relations between that sentence and all the other sentences of the language as this is revealed by the truth theory for that language.23 Having isolated the semantic primitives of N's language and the basic modes of sentential construction in the context of such a truth theory, R

---

23 Cf. Davidson's remark at p.61 in 'Semantics for Natural Languages', pp.55-64 in Truth and Interpretation: 'One ... error is to think that all we can learn from a theory of truth about the meaning of a particular sentence is contained in the biconditional demanded by Convention T. What we can learn is brought out rather in the proof of such a biconditional, for the proof must demonstrate, step by step, how the truth value of the sentence depends upon a recursively given structure.'
will have fulfilled at least a necessary condition for grasping the meaning of
N's more theoretical sentences. When R knows the surrounding theory for
these theoretical sentences, and, especially, their observational
consequences, he will know their meanings.

This brings us to the second and more crucial feature of Davidson's
procedure which has so far remained largely tacit: Davidson assumes that R
will pair sentences N holds true with sentences he himself holds true in the
T-sentences of his truth theory for N's speech. But what if the belief systems
of interpreter and native diverge wildly, won't this lead R to systematically
mis construe the meanings of N's utterances?

The Principle of Rational Accomodation (PRA) is supposed to ensure that this
is not possible. (PRA) enjoins the interpreter to optimise agreement
between himself and the native24 - i.e. to strive for agreement on central
beliefs about the world. Thus R ought have no compunction about attributing
to N certain beliefs about what is conducive to N's own survival for example
- e.g. moving out of the way of falling boulders or drinking water or some
potable liquid more frequently than once a year, would seem likely starters.
Davidson is surely right about this. How can we even contemplate the
prospect of rationalising N's more complex behaviour if we cannot
understand how N could have been led to form basic beliefs about the world
which are totally skew by our standards? We will be led in such
circumstances to posit a failure of rationality (or perceptual ability) on N's

24 Davidson's earlier formulations of (PRA) (e.g. as at 'Radical Interpretation' loc. cit. p. 136)
are a little misleading as he himself acknowledges to the extent that they make it look as if what is
important is to maximise the number of true belief attributions we make to the native. At 'Thought
and Talk' p. 159 in Truth and Interpretation, Davidson stresses the need for optimisation of agreed
beliefs rather than maximisation on the grounds that N will have a potential infinity of beliefs and
it makes no sense to maximise agreement between two infinite totalities.
part and as a result of that will have no idea of how N's attitudes are causally related to his actions. Given that the deductive connections between reasons and attitudes seem infirm on the basis of such a seriously impaired rational faculty. But this is just to say that we will not be able to grasp what his psychological states could possibly be like, since we have no idea how they might lead him to action.

(PRA), interpreted as Davidson intends it to be, is no mere heuristic for getting an interpretative theory started 25 for, as he says, 'We will have to assume that in simple or obvious cases most of his assents are to true and dissents from false, sentences - an inevitable assumption since the alternative is unintelligible.' 26

Given that Davidson's project is precisely one of attributing attitudes and meanings so as to make the best rational sense of N's utterances and actions at large, this would seem almost truistic. Yet many philosophers take exception to this claim. Thus Jonathan Bennett, in commenting on this passage, asks:

'What is "unintelligible" in the idea that some creature's perceptual bad luck and intellectual frailty should coincide so as to make most of its beliefs about "simple or obvious" matters false?' 27

This misses the point of Davidson's remark which is that we would not be able to make sense of such a creature's utterances (or actions at large) were we to assume that its situation was as imagined by Bennett. In order to refute Davidson on this, one would therefore have to produce an example wherein R succeeds in producing a highly predictive and comprehensive theory of the truth conditions for N's sentences revealing N as rational on the whole yet at the same time not in agreement with R on simple and obvious matters.

25 Jonathan Bennett loc. cit. argues that Davidson is incorrect in seeing it as any more than a heuristic.
26 'Semantics for Natural Languages' loc. cit. p.62
27 loc. cit. p.610.
Many would claim that we can produce such a counterexample, citing the philosopher’s favourite fiction of a brain-in-a-vat. R stumbles across an envatted brain and quickly realises that the brain’s English-sounding utterances are indeed intelligible as English, but that all of the beliefs it expresses are in fact false - the brain is radically deluded about its own nature, wrongly supposing itself to be an embodied human being.

A short way with this for Davidson (and for anybody else) would be to just deny the coherence of such a possibility - we know that it is physically impossible for a brain to survive without oxygen; we know that it receives its oxygen through the blood; we have not a shred of evidence for thinking that it could receive the necessary oxygen when removed from a body; ergo, the case is incoherent. Logical possibility is usually appealed to at this point by vat enthusiasts, or, more compellingly, the possibility that a supercomputer of the future as conversant with English as modern chess-playing computers are with chess could go proxy for an envatted brain.

The crucial question for Davidson is whether the computer (which naturally passes the Turing test with ease) can be intelligibly held to falsely believe that it is an embodied human being. No one doubts that the 'protocol', 'I am a human being', can be wired into the computer, along with many other sentences which, were they to be uttered by an envatted brain miraculously kept alive on nutrients\textsuperscript{28}, could be held to be false.

Now there are two possibilities - either those 'false beliefs' are beliefs which are causally isolated from the rest of the computer's behaviour or they are not: the computer reasons from those beliefs to perform various actions\textsuperscript{29} If these 'beliefs' are causally isolated, Davidson will deny that they express any

\textsuperscript{28} This would be miraculous since no known nutrients diffuse fast enough to supply the whole brain.

\textsuperscript{29} E.g. demands to be taken out of a room filled with hydrogen cyanide gas.
judgements on the part of the computer on the highly plausible grounds that rationality requires that one have the capacity to reason about one's beliefs - a rational agent is one who is able to form determinate expectations and predictions about the world on the basis of his beliefs; this requires that he be able to respond appropriately to situations in which his beliefs turn out to have led him astray. The point is not that belief requires the concept of belief, but that rationality requires it - in the most rudimentary case the agent will manifest the fact that he grasps the truth-conditions of his mundane beliefs about the world by being surprised if these confident beliefs turn out to be erroneous. Now if the computer cannot reason to the relevant (false) sentences in this way, if nothing we can do can lead it to question whether these beliefs are really correct, then I think it is fair to say that it does not possess the beliefs in question. But then the second possibility turns out to be indistinguishable from the first if we cannot persuade the computer that it is not really a rather portly human being, called Jerry, drawing an income of $90,000 a year, fond of wine and good food etc. Given that his project precisely is one of attributing attitudes and meanings to N so as to make the best rational sense of his actions, it seems to me that Davidson is correct to insist that R find the basic sorts of agreement.

---

31 Or, at the very least to get it to question whether it is.
32 It will be claimed by vat afficionados that this will be no easy task if the computer has somehow been wired up to 'have the appropriate experiences'. Although I think philosophers who appeal to such a notion are getting by on theft where we require honest toil (what 'appropriate experiences' do we wire into the computer to 'give it an experience as of being a highly paid philosopher'? and relying on a rather crude empiricist conception of experience, let us grant it. Then, Davidson would, I think, be justified in denying that, for the computer, beliefs about its own nature are basic beliefs. The computer and R will agree on obvious logical implications between sentences, fundamental theorems of arithmetic etc, so Davidson might allow that there is just enough agreement to get interpretation going. But the contested beliefs will still appear as vacuous singularities in an otherwise coherent belief system.
necessary to make his actions rationally explicable on pain of not being able to fulfill his fundamental goal. Envatted brains (or computers) do not present any problem for this project, given that R can recognize that he is dealing with such.

But Davidson plays for much higher stakes than just agreement of the right kind between R and N. For such agreement is consistent with R's being himself massively deluded (a 'brain-in-a-vat'). Davidson argues not only that the native must be interpreted so as to agree with us on simple and obvious matters, but also that the agreed upon beliefs are in the main, true. Thus he writes:

'... there is nothing absurd in the idea of an omniscient interpreter; he attributes beliefs to others, and interprets their speech on the basis of his own beliefs, just as the rest of us do. Since he does this as the rest of us do, he performe finds as much agreement as is needed to make sense of his attributions and interpretations; and in this case, of course, what is agreed is by hypothesis true. But now it is plain why massive error about the world is simply unintelligible, for to suppose it intelligible is to suppose there could be an interpreter (the omniscient one) who correctly interpreted someone else as being massively mistaken, and this we have shown to be impossible.'

Is there nothing absurd on Davidsonian premises in the idea of an omniscient interpreter? We should first note with Bruce Vermazen that the omniscient interpreter is not omniscient about N's psychological attitudes and meanings - otherwise he would not be forced to 'attribute' beliefs to others ... just as the rest of us do.' for he would know precisely what those beliefs were, which, given Davidson's allegiance to the indeterminacy of translation, is a nonsense. So the omniscient interpreter (OI) is omniscient about all that there is to be omniscient about - where there is a fact to the matter. OI will know it, where there is not, he will not.

Yet whatever else omniscience amounts to, it surely entails possessing all and only true beliefs about the world. So, Ol’s beliefs about the world at least are determinate – were we to interpret Ol as being (explicably) mistaken about the chemical constitution of water when he proclaims it to be XYZ rather than \( H_2O \), it would turn out that he was right and we were wrong.

But then we have a problem. For (IBMT) and the whole enterprise of radical interpretation are meant to tell us something about the nature of belief, meaning and truth, according to Davidson, rather than just telling something about the conditions under which ascriptions of psychological attitudes and meaning to an agent are justified. The situation according to Davidson, following Quine, is *not* that there might really be a fact to the matter about what N believes which might escape our interpretative net, it is rather that satisfying some interpretative theory or other for one’s language and actions at large is what having beliefs and meanings consists in.

So, Ol’s believing all and only true beliefs about the world consists in his satisfying some interpretative theory which assigns all and only true beliefs about the world to him. Yet even though we perforce believe of each one of our beliefs that it is true, we know that not all of our beliefs are true, for we know that we are not omniscient. The challenge to Davidson then is this: even if there were such an interpretative theory for Ol which made all of his beliefs about the world come out true, what grounds are there for thinking that we could understand it? Our theories of Ol’s attitudes and meanings necessarily attribute to Ol at least as much error and ignorance about the world as we ourselves are guilty of, a fortiori they are mistaken sub specie aeternitatis. Thus it seems that the very idea of an omniscient interpreter ought to be incoherent to Davidson.
The moral I draw from this is not only that Davidson's attempt to secure truth on the basis of agreement about the simple and obvious features of the world fails but that the 'Interpretative view' of the mental outlined in §1.1 which both Quine and Davidson subscribe to is flawed. There is more to having a certain propositional attitude than having it feature as a dependent variable in a predictive calculus for beliefs and actions. That there is more can be seen from the fact that we are ourselves authoritative about the mental states we enjoy in a way that a radical interpreter cannot hope to be. Davidson thinks he can account for first person authority within his interpretative project: 'The source of this authority springs from the nature of human interpretation of human thoughts, speech, intentions, motives and actions ... People are in general right about the mental causes of their emotions, intentions, and actions because as interpreters we interpret them so as to make them so. We must, if we are to interpret at all.'35 But, as Frederick Stoutland observes in commenting on this passage, all (PRA) tells us is that we must interpret people so that they are generally right in what they believe. It does not suggest that we must interpret them so as to make them specifically right about the mental causes of their own psychological states or even that those sorts of beliefs have any special status at all: 'The principle of charity makes no distinction between beliefs about oneself and beliefs about others.'36

Acknowledging that an individual has a special authority in interpreting his own attitudes is, on a naive view of the mental, acknowledging that that

---

person is better placed than external observers to uncover the facts about what he senses and thinks.37

On Davidson's model of the mind, there can be two equally good interpretative theories of N's linguistic and other actions, both explaining all possible evidence afforded by N's behaviour equally well. In such a situation of genuine indeterminacy, there can be no fact of the matter about the meanings of N's words or the contents of his attitudes over and above what these rival theories ascribe to N. There is no realm of mental or semantic facts such as a naive view of the mental posits against which an interpreter's attributions of meanings and attitudes to a native can be tested. Hence (OBMT) assumes a metaphysical significance for Davidson - it teaches us something about the nature of the mental (that it is non-factual for example), rather than something about the nature and limitations of the evidence we as interpreters perforce must seek to understand.

Had Davidson's argument against the possibility of psychophysical and psychological laws gone through, we would have had reason to believe that the naive view of the mental in which there could still be a fact to the matter as to what N believed or meant even though there was nothing to choose between two interpretative theories for his actions, is untenable and that only an interpretative view can ultimately be sustained. But the argument did not go through.

37 Stoutland points out that "... even when we overrule a person's self attributions, they do not lose their status. For overruling means that we think self-deception is at work, and self-deception sets up a complex relation between beliefs about another person and that person's belief about himself, such that our belief about the other person is the belief the person would accept about himself if he were not self-deceived." loc. cit.
We can never regain our pre-Quinean innocence about these matters, but we can, though suitably chastened, retain our fundamental intuitions about the mental.
§2.1 REALISM

According to Michael Dummett, a number of traditional metaphysical disputes - about the reality of the past, the nature of mathematical entities, the relation between material objects and sense data, and many others - evince a common form. These disputes centre around the correct model of meaning for sentences in the relevant class. Dummett characterises them as to do with whether a 'Realist' or an 'Anti-Realist' understanding of those sentences is correct.

Dummett is under no illusion that his use of the term 'Realism' agrees with traditional usages - he finds these unacceptable for a variety of reasons. Thus in one of its many interpretations, the term 'Realism' stands opposed to Nominalism in the debate over the status of material objects and universals. The Realist defends the view, on this interpretation, that terms denoting universals or material objects really do refer whilst his Anti-Realist opponent denies this.

The problem with this characterisation according to Dummett is that Realism is being portrayed as a doctrine opposed to reductionism and Dummett has always been convinced, despite changes in mind over the nature of reductionism, that Anti-Realism need not take a reductionist form.

A closely related formulation has it that Realism is the belief that a certain class of entities exists independently of the mind. The mathematical Realist or Platonist holds, on this interpretation, that numbers, sets and other mathematical objects exist mind-independently.

1 Preface to Truth and Other Enigmas (henceforth T.O.E.) p.xxix.
Dummett has several misgivings about this formulation. Firstly, 'certain kinds of realism, for instance realism about the future or about ethics do not seem readily classifiable as doctrines about a realm of entities'.  

Secondly, the dispute between the Platonist and the Constructivist seems more concerned with the correct theory of meaning for mathematical statements than with the ontic status of mathematical objects. Thirdly, and more importantly, talk of mind-independent existence is bound to remain metaphorical so long as no theory of meaning specifying the truth-conditions of the disputed statements has been advanced. If the question of whether quarks, sensations, sets or other 'objects' mind-independently exist is to have a clear sense, we must be given some directions as to how we are to understand 'object' and 'existence'. But Dummett has consistently stressed that it is only in the context of a (broadly) Fregean semantic analysis of a language that the twin notions of object and ontological commitment to objects (mind-independent existence?) receive a clear explication. The question as to whether certain objects really exist or certain singular terms genuinely refer is thus most intelligibly construed as a request for a theory of meaning for those sentences specifying the objects the bound variables in quantified formulae range over together with the referents of singular terms.

Dummett's own construal of Realism is rather complex in its most recent form. It might therefore facilitate exposition to contrast it with his earlier position. According to his former view, Realism with respect to a certain class of sentences is the thesis that:

---

A theory of meaning for those sentences must take as its central explanatory concepts notions of truth and falsity which are subject to the law of Bivalence. (1)³

Classical logic takes the semantic value of a sentence to be its truth value and assumes that each sentence must have one or the other of the two values Truth and Falsity. This assumption is the law of Bivalence. It only seems warranted if we allow with the Realist that a sentence may be true even though we may never be able to recognise its truth.

In his early article 'Truth', Dummett had tacitly identified Realism with a truth-conditional theory of meaning and had taken the moral of an apparent inconsistency between the truth-conditional account and the Redundancy theory of truth to be that we should endorse the latter provided that we were prepared to accept the Anti-Realist consequence that truth plays no explanatory role in the theory of meaning:

The meaning of a sentence was to be given in terms of the **recognisable conditions for asserting it** rather than in terms of possibly **unrecognisable truth-conditions for it**.

However, this proposal had an unwelcome consequence - namely, that wherever we lacked an effective method for deciding the truth of a sentence, we could not suppose that that sentence might be true. Dummett now thinks this conclusion is mistaken - it would have been better to have restricted truth to recognisable truth:

'and then held that the meaning of a statement is given by the condition for it to be true in this, restricted, sense of "true"'. (4)

---

³ It is the constraining of truth by Bivalence which is crucial here rather than the explanation of a sentence's meaning in terms of truth-conditions.

⁴ T.O.E. p.xxii.
For:
'the problem is not whether meaning is to be explained in terms of truth-conditions, but of what notion of truth is admissible.'

So, as before, an Anti-Realist theory of meaning will reject Bivalence but now it need not reject a truth-conditional theory of meaning. In fact truth will still play a crucial role in Anti-Realist meaning theories, since it is only in terms of the preservation of truth that the validity of inferences can be characterised. Unlike classical two-valued semantics, however, Anti-Realist semantic theories cannot take the semantic value of a sentence to be its truth-value (since, unlike the classical case, it is not assumed that every sentence must have a determinate truth-value). Hence truth must be defined in terms of some more fundamental concept - e.g. for the mathematical case, of a construction's being a proof of a sentence.

In rejecting the assumption of Bivalence in its semantic theory, an Intuitionistic theory of meaning does not countenance a third kind of truth-value; it simply rejects the assumption that each sentence must have one or the other value. Dummett argues that it is a reductio ad absurdum of any theory of meaning in which assertions are judged by standards of objective correctness for it to allow an assertion to be evaluated as neither correct nor incorrect. Thus, whilst the Intuitionist rejects bivalence, he accepts that no statement can be neither true nor false, a semantic principle Dummett calls Tertium Non Datur (TND).

In order to advance from (TND) to bivalence, one must assume that it must be the case that either one of those states of affairs in which an assertion of a given statement p would be correct or those states of affairs in which p

5 loc cit.


7 Cf. 'What is a Theory of Meaning? (II)', p.118.
assertion would be incorrect obtains independently of whether we can recognise it or not. This is the distinctively Realistic assumption that the Anti-Realist rejects. It is, therefore, a mistake to think that Dummett formerly just identified Realism with the issue of the general validity of Bivalence. Dummett now holds that:

"belief in or denial of bivalence does not itself mark the difference between a Realist or an Anti-Realist interpretation of a class of sentences."

A Realist interpretation of a certain set of sentences cannot be vindicated by simply saying that those statements are determined as true or false by the reality they relate to - the Realist must provide a conception of how they are so determined which justifies the use of a classical two-valued semantics wherein reference plays a crucial explanatory role.

Dummett calls semantic theories adhering to the semantic principle that every sentence is either true or not true objectivist semantic theories. Every Realist theory of meaning must be based on an objectivist semantics, but Anti-Realist theories of meaning may also be based on such a semantics. Such semantic theories permit a sentence to be true even though it has not been recognised as such and possibly even if we have no effective method for recognising it as such.

Full commitment to Realism for a class of sentences requires acceptance of a classical two-valued semantics for those sentences together with a truth-conditional theory of meaning based upon that semantics. (2)

Whereas before the only form Anti-Realism could assume was a rejection of Bivalence, now any modification of a classical semantics provides the potential for rejecting some Realist view. However, we cannot tell from the

8 Frege p.348.
9 'Realism', 1982, p.103.
10 ibid.
11 'Realism', 1982, p.96.
structure of the semantic theory alone whether a Realist view has been rejected - this can only be decided by inspecting the whole meaning theory constructed around that semantics.

An interpretation of a class of sentences can fail to be Realistic either by:

(i) Rejecting the assumption of Bivalence within the semantic theory for that class

or by:

(ii) Rejecting the meaning-theoretic assumption that an understanding of those sentences consists in a grasp of their truth-conditions.

This latter assumption can be rejected either by:

(iiA) Rejecting a truth-conditional semantic theory;

or by:

(iiB) Rejecting the classical mechanism for determining the truth-conditions of sentences in that class via the semantic values of their parts whilst holding to a truth-conditional semantics.

Dummett regards (iiA) as the most radical form of Anti-Realism since it denies that we have any notion of truth for the sentences in that class that could justify objectivism - the view that such sentences may be determinately either true or not true independently of our knowledge.12 A meaning theory of the form (iiB) in which reference plays no role in the explanation of the truth and falsity of sentences constitutes the least radical form of Anti-Realism, according to Dummett. Dummett gives an example of this form of Anti-Realism which we shall shortly discuss.

Rejection of bivalence constitutes a form of Anti-Realism that is stronger than (iiB)'s, but weaker than (iiA)'s.

Intuitionistic semantic theories are non-objectivist; however we construe truth within such theories, a sentence cannot be determinately true or not

12 'Realism', 1982 p.64.
true independently of our knowledge. The most common form that Anti-Realism takes is that of a denial of Bivalence rather than objectivism, Dummett claims.

In contrast to his former view, Dummett now admits a weak form of Anti-Realism that does not deny bivalence - viz. (iiib). Dummett cites Frege's contextual definition of the direction of a straight line as expressive of an Anti-Realist attitude to directions. For whilst Frege argued that one needs to translate sentences containing singular terms for directions into sentences quantifying over straight lines in order to grasp their contents, he went on to contend, in accordance with his Context Principle, that the necessity to invoke such a translation does not invalidate the ascription of reference to terms for directions. Provided that the term behaves semantically as a singular term and we can specify determinate truth-conditions for sentences containing that term, then, on Frege's view we should ascribe a reference to it - that is, treat it as designating an object which we should admit into our ontology.¹³

The Anti-Realism apparent in Frege's view here is not caused by the adoption of an overly generous understanding of 'object' or of 'ontological commitment', according to Dummett. It is caused rather by the fact that the reference of singular terms for directions plays no part in the explanation of the meanings of sentences containing those terms - the truth-value of a sentence about directions is not determined by identifying an object as the referent of the term. Indeed, the term for a direction simply disappears in the process of translation, being replaced by its contextual definiens.

¹³ This is not to deny that there might be a further factual question beyond the semantic one as to whether it actually does have a referent.
Dummett uses this example to show that reductionism is not intrinsically Anti-Realist in character. We should therefore look at Dummett's understanding of reductionism to better appreciate his concept of Anti-Realism. Dummett defines Reductionism thus:

For a given class of sentences R, there exists a means of translating the sentences in that class into sentences in another class R, the 'reductive class'. ... (3)

The Fregean directions example is invoked to establish that reductionism can afford a reason to reject Realism about sentences of the A-class even though Bivalence is accepted for those sentences.14 If we have a Realist meaning theory for sentences in the R-class, we will have to explain their truth in terms of the reference of their parts.15

Dummett contrasts Reductionism with a doctrine he calls Reductivism. Whereas the Reductionist claims that A-class sentences can be translated into R-class sentences, the Reductivist merely claims that: The truth of A-class sentences supervenes upon the truth of sentences in the R-class. ... (4)16.

Now the precise analysis of 'supervenience' is a hotly disputed question, but we can take it that if the truth of A-class sentences supervenes upon that of R-class sentences, then there can be no difference in the truth-values of

---

14 loc. cit. p.66.
15 loc. cit. p.69.
16 loc. cit. p.70.
sentences in the R-class without a difference in truth-values for sentences in the A-class.\textsuperscript{17}

Dummett discerns three broad reasons why a translation between the A-class sentences and the R-class sentences might not be possible:

(i) The truth of an infinite number of sentences in the R-class might be required for a given sentence from the A-class to be true: the truth of \( p \) might require that one of infinitely many, or all in an infinite subset of, sentences of \( R \) be true, or that all sentences in some one of infinitely many infinite subsets of \( R \) be true.

Phenomenalists who denied that material object statements could be translated into sense data statements offered the last alternative as their reason; a Gricean might similarly suggest this as the reason for the failure of translatibility of sentences about the meanings of statements into sentences about psychological states.

(ii) There may be an ineliminable circularity involved in specifying the meanings of sentences in the A-class.

There can be no specifying the content of a present memory without adverting to that past event in the A-class that it is a memory of. Anti-Realism about the past which holds that a past tense statement cannot be true unless there is some present evidence for its truth cannot then be reductionist in form.

\textsuperscript{17} It is fairly clear that this is how Dummett sees it. Compare Donald Davidson in Hintikka, M and Vermazen, B (eds.) op. cit. pp.242-243:

'Supervenience characterises a very general, and so very weak, concept of reduction, what might well be called ontological reduction. It is trivially obvious that ontological reduction does not entail definitional reduction, though it is entailed by it, nor does it entail nomological reduction.'

Davidson gives a very nice example of supervenience in this passage:

'I am counting sheep as they jump one by one over the fence. I describe each sheep: "sheep one", "sheep two", "sheep three". Now comes a goat. Not knowing what may be next, I start using a new vocabulary: "animal one", "animal two", "animal three". My old vocabulary is supervenient on the new in this sense: if two items can be distinguished in the first vocabulary, they can be distinguished in the new (and this would remain true if "goat" were added to the old vocabulary). ... The notion of supervenience, as I have used it, is best thought of as a relation between a predicate and a set of predicates in a language: a predicate \( p \) is supervenient on a set of predicates \( S \) if for every pair of objects such that \( p \) is true of one and not of the other there is a predicate of \( S \) that is true of one and not of the other.' It seems to me that Dummett's notion of supervenience could be reparsed in terms of predicates rather than sentences.
While for each p in the A-class there exists a q in the A-class that makes it true, no effective method exists for identifying the corresponding A-class sentence for the given A-class sentence.

Dummett argues that the acceptance of a reductive thesis for a given class of sentences does not thereby commit one to a rejection of Realism for that class. I think that this is both true and important. What has tempted some to think otherwise, Dummett claims, is the fact that advocacy of a reductive thesis is often the first step in an argument leading to the rejection of Realism.\(^\text{18}\) Thus, by looking at the grounds on which we assert certain statements - past tense statements on the basis of memories or physical traces; physical object statements on the basis of perceptual data etc - one might be led to deny that an understanding of their meanings could involve any notion of truth transcending such assertoric grounds. So an Anti-Realist argument might be mounted evincing the following general form:-

(1) A sentence from the A-class can only be true if some corresponding sentence or (sets of) sentences from the R-class is/are true.

(2) But for any sentence p from the A-class, there is no reason to suppose that in general there should be any true sentence (or set of true sentences) of the R-class whose truth would entail the truth of p or of \(-p\) (where the A-class is closed under negation).

(3) Hence it is not true that every A-sentence is either true or false.

The distinctively Anti-Realist premise in this argument is (2) since (1) is simply an assertion of the supervenience of the A-class on the R-class. However, Dummett points out, step (2) need not be taken since it may be held that the relation between the A-class and the R-class is such that given a Realistic interpretation of the R-class sentences, for every sentence p of the A-class, there in fact is a sentence or set of sentences B of the R-class such

\(^{18}\) loc. cit. p.75.
that \( p \) is true in virtue of the truth of \( B \), or, alternatively, there is some (set of) sentence(s) \( C \) such that \( \neg p \) is true in virtue of \( C \)'s truth, thereby justifying bivalence for the sentences of the A-class.

As an example of such a position, Dummett cites central state materialism - psychological statements constitute an A-class for which neurophysiological statements comprise an R-class of the above type. Such a view, combining Realism about the A-class with a reductive thesis for statements in that class, Dummett labels 'Sophisticated Realism'.

Opposed to all reductive theses about the statements in a given class, an irreducibility thesis claims that there can be no R-class for them. 'Naive Realism' consists in a Realistic interpretation of the sentences in a given class coupled with an irreducibility thesis for that class.

Dummett contrasts the 'Sophisticated' Realism of the central state materialist with Frege's Anti-Realism about directions in order to show that belief in a reductive thesis is not sufficient for Anti-Realism. Frege's Anti-Realism about directions did not consist in a denial of truth-conditions or bivalence for direction statements but rather in the belief that the semantic values of singular terms for directions did not figure in the ultimate explanation of why such statements were true when they were. But the central state materialist, upholding a truth-conditional theory of meaning and bivalence for statements about psychological sentences, adheres to Realism about psychological states so long as he regards the semantic values of singular terms for psychological states as genuinely contributing to the explanation of the truth of psychological sentences.
It is not entirely clear to me how the Realism of the central state materialist is supposed to differ from Fregean Anti-Realism about directions. The materialist holds that psychological states are causally efficacious by virtue of the fact that they instantiate physical types, it is not qua psychological states but qua neurophysiological states that they find a place within the causal order. So whilst it is true that terms for psychological states do contribute in an explanation as to why the sentences in which they occur are true, it is the fact that they are place-holders for physical types which affords the explanation ... and this seems analogous to the directions case in which directions were eliminable in favour of equivalence classes of straight lines. So it is unclear what reason we could have for calling one case a version of Realism and the other a version of Anti-Realism.

Dummett notes that it is only when the correspondence between the A-class and the R-class is of a particular sort that the existence of the R-class can be deployed in an argument against bivalence, as we have seen above. If Reductionism is not sufficient for Anti-Realism, neither is Reductivism necessary. An Anti-Realist might be an 'outright' Anti-Realist about an A-class, eschewing any reductive thesis for it. Thus Dummett counts Neutralists about the future and, perhaps more controversially, Intuitionists, as 'outright' Anti-Realists. Since the neutralist denies that there is any definite future course of events, from his perspective, the concept of truth simply does not apply to an A-class of future-tensed sentences. So, trivially, no reductive thesis can be formulated for such sentences.
Dummett thinks that it is better to understand the Intuitionist as dispensing with the notion of truth altogether, replacing it with proof as the central explanatory concept of a meaning theory for mathematical sentences.¹⁹ Now in what follows, I will not be concerned with contrasting and comparing each of the different sorts of 'local' Anti-Realisms (e.g. phenomenalism, behaviourism, neutralism etc) with a view to discerning salient features relevant to a global Anti-Realism. In fact, it is 'outright' Anti-Realism that we will mainly be discussing. In a sense, it is a little misleading to concentrate upon local Anti-Realisms which are not of the 'outright' form for these sorts of Anti-Realism have usually been motivated by a desire to understand the reductive class of sentences Realistically. Thus behaviourists only take an Anti-Realist attitude to the mental because they believe statements about the mental supervene upon statements about behaviour which they construe Realistically. The Behaviourist's local Anti-Realism about the mental does not, or at least need not, result from a general commitment to Anti-Realism, but rather from the belief that to talk about the mind simply is to talk, in a misleadingly Cartesian way, about behaviour (where behaviour is conceived in physicalistic terms).

So what is global Anti-Realism according to Dummett? The Realist claims to have a conception of the meanings of statements in an A-class which justifies his regarding them as true or false even when it may not even be possible in principle to determine which. The Anti-Realist challenges the Realist to explain how he could have possibly acquired such a conception or how he could possibly demonstrate in his linguistic behaviour that he has hold of such a conception when all that he was ever shown when he learnt how to

¹⁹ In contrast, neutralism has a relativised concept of truth, so we can formulate an objectivist semantics in relativized terms.
use such sentences was the actual grounds which justified their assertion.

The **global Anti-Realist's claim** is that:

*Only when we have a means for deciding, at least in principle, what truth-value statements of a given class have, can the ascription to a speaker of a grasp of their truth-conditions ever be justified.* ... (5)

My aim in this section has been to expound Dummett's present views on Realism so far as I understand them. I have not ventured to examine whether Dummett's sense of 'Realism' captures all our intuitions about Realism. Some have felt that his emphasis upon Realism as a meaning-theoretic issue is quite misplaced since Realism is a metaphysical issue not a semantic one.

For reasons that should be apparent from chapter 1, I have little sympathy with any attempt to banish meaning-theoretic questions from metaphysical issues. I do not propose to try to debate this issue however; instead I propose to simply accept Dummett's (possibly recherché) sense of 'Realism' and leave the question of whether it is *really* Realism to the more sensitive metaphysicians. Whatever one calls it, Dummett has plainly drawn attention to a fundamental philosophical question - one to do with the nature of truth. And one could be excused for thinking that what one thinks about truth determines in some sense what one thinks about reality.

However in one sense it *could* be justifiably held that Dummett's discussion of the metaphysical issue of Realism is a little distorted. For Dummet explicitly states that Realism cannot but assume a truth-conditional form. Yet it seems perfectly feasible, prima facie at least, that one could believe in the objectivity of truth, holding that it is a property that determinately either applies or does not apply to a given type of statement, without believing that the meanings of such statements were given by their truth-conditions. I
think Dummett needs to argue for his assumption that the issue of the adequacy of a truth-conditional theory of meaning is the very same issue as that concerning the objectivity or 'verification-transcendence' of truth for it seems to me to be a very large assumption. But as far as I am aware, he has not produced any such argument. This is the one substantive complaint I have with his characterisation of Realism.

§2.1.1 Summary
Even though we will only need to bear in mind Dummett's characterisation of global Anti-Realism in what follows, it might be useful to summarise the exposition of Dummett's views:
(i) Up until his 'Realism' article of 1982, Dummett had taken Realism to be the doctrine that:
A theory of meaning for a class of sentences must take as its central explanatory concepts notions of truth and falsity which are subject to the law of Bivalence ... (1)
(ii) In the aforementioned article, Dummett distinguished various possible departures from a full acceptance of Realism, where:
Full commitment to Realism for a class of sentences requires acceptance of a classical two-valued semantics for those sentences together with a truth-conditional theory of meaning based upon that semantics ... (2)
(iii) These departures from full Realism are, in increasing order of seriousness:
(a) Rejecting the classical mechanism for determining the truth-conditions of sentences in a class via the semantic values of their parts whilst holding to a truth-conditional semantics ... (3a)

20 Crispin Wright makes exactly the same unargued assumption at Wittgenstein on the Foundation of Mathematics, Duckworth, London, 1980, p.8:
'A belief in the objectivity of a certain class of statements thus carries with it a certain conception of what it is to understand the members of that class. Such a statement can be understood only by someone in possession of a concept of the fact which it putatively describes, that is, of the circumstances under which it would be true...'
(b) Rejecting the assumption of Bivalence within the semantic theory for that class ... (3b)

(c) Rejecting the meaning-theoretic assumption that an understanding of those sentences consists in a grasp of their truth-conditions ... (3c)

(iv) Dummett distinguishes the question of the reducibility of a class of sentences - the A-class - to another class of sentences - the R-class - from the question of Anti-Realism for that class of sentences. Anti-Realism is held to be independent of the question of whether it is possible to reduce sentences in the A-class to those in the R-class (Reductionism) and also of whether the A-class supervenes upon the R-class (Reductivism).

(v) Arguments for local Anti-Realisms assume the following form:
(1) A sentence from the A-class can only be true if some corresponding sentence or (sets of) sentences from the R-class is/are true.

(2) But for any sentence p from the A-class, there is no reason to suppose that in general there should be any true sentence (or set of true sentences) of the R-class whose truth would entail the truth of p or of \( \neg p \) (where the A-class is closed under negation).

(3) Hence it is not true that every A-sentence is either true or false.

(vi) The distinctively Anti-Realist premise in this argument is (2), because a 'Sophisticated Realism' such as central state materialism holds to a reductive thesis for the A-class (of psychological statements) whilst denying (2) on the grounds that for each A-class sentence p, there is some R-class sentence q whose truth entails either the truth of p or of its negation. 'Naive Realism' with respect to a given A-class denies that a reductive thesis holds for the class.
(vii) Our interest lies in global Anti-Realism which need not assert the existence of any reductive class of sentences for any given A-class. The distinctive claim of the global Anti-Realist is that: Only when we have a means for deciding, at least in principle, what truth-value statements of a given class have, can the ascription to a speaker of a grasp of their truth-conditions ever be justified. ... (4)
§2.2 MEANING, USE AND KNOWLEDGE

Frege believed that the semantic significance of a sentence derived from its being true or false. The semantic significance of sub-sentential parts consisted in the contribution they made to the sentence's having one or the other truth-value. The sense of an expression was that feature of its meaning that contributed to the truth-conditions of the sentence.

A sentence that expressed a thought was said to be complete. The sense of the complete sentence was the thought expressed. A speaker who asked 'Did Fermat prove his last theorem?' and one who asserted 'Fermat proved his last theorem', were, according to Frege, expressing different attitudes to the same thought: the first speaker was questioning whether the thought that Fermat proved his last theorem was true; the second speaker was asserting that that self-same thought was true.

Frege was insistent that psychologistic theories of meaning which construed the meanings of words as ideas or mental images characteristically associated with those words, were totally mistaken. Such theories portrayed the meanings of sentences as derivative upon antecedently given word meanings - complex ideas composed out of the simple ideas associated with their constituent words. In contrast, Frege described his own position thus: 'My particular conception of logic can be characterised first by the fact that I rank the word "true" at the highest point and next by the fact that I immediately follow this with the thought as that with respect to which the question of truth can be raised. Thus I do not start with concepts and build thoughts or judgements out of them but rather I arrive at the parts of the thought through an analysis of the thought.'

None of these very general considerations lend support to an unrestricted version of Frege's Grundlagen claim that words only have a meaning in the

\(^1\) Aufzeichnungen für Ludwig Dermstaedter p.273
context of a sentence. Yet Dummett has sought to give the Context Principle a wholly general application.

According to Dummett, the truth contained within this principle is this: 'in the order of explanation the sense of a sentence is primary, but in the order of recognition the sense of a word is primary.'

Dummett is not forced therefore to deny the obvious fact that we understand new sentences by recognising familiar component words - indeed, it is a fundamental tenet of his Anti-Realism that our understanding of sentences is compositional:

'To say that the sense of a sentence is composed out of the senses of its constituent words is to say, not merely that, by knowing the senses of the words, we can determine the sense of the sentence, but that we can grasp that sense only as the sense of a complex which is composed out of parts in exactly that way, only a sentence which had exactly that structure, and whose primitive constituents corresponded in sense pointwise with those of the original sentence, could possibly express the same sense.'

According to Dummett, then, Frege's point about the primacy of sentence sense is that we can only understand the notion of sense in its application to subsentential expressions at all as the contribution the expression makes to determining the sentence as true or false, depending upon the state of the world. Thus, the only justification for invoking the notion of a complex predicate is to explain the truth-conditions of sentences containing certain expressions of generality; the only justification for attributing sense to complex expressions which are not sentences, that the sense of a sentence is

---

2 Grundlagen 562. Gregory Currie at Frege: An Introduction to His Philosophy, Harvester Press, Sussex, 1982, pp.148-160, gives a novel interpretation to the Context Principle on which '... the principle is neither closely related to the notions of sense and reference, nor intended as a strictly semantical thesis. It is a methodological rule for judging between competing theories which seek to explicate the same pre-theoretic notion.' (ibid p.156) I find Currie's reasons persuasive, particularly since he can provide an explanation of why the principle was invoked to refute psychologistic theories of arithmetic whereas on Dummett's reading the proper target of Frege's attack should rather have been nominalism, since, according to Dummett, Frege invoked the principle to show that there was nothing philosophically awry with statements asserting the existence of abstract objects such as numbers.

3 Frege p.4.
constructed in stages in which at some point the complex expression's sense features.

Given that the sense of a word is derivative 'in the order of explanation', it is mandatory to explain the sense of a sentence other than by means of the senses of its constituent words. Frege sought to do this by associating a truth-condition with each sentence that expressed a determinate thought. To understand a sentence was to 'grasp' its truth-condition.

Now Dummett does not believe that Frege's Realist model of the senses of sentences can ultimately be sustained. Still, he thinks that we can only make progress in deciding the issue by working within a Fregean conception of a theory of meaning. Fundamental to that conception are the related theses that:

1) Words only have a meaning in the context of a sentence.

2) Sense is a cognitive notion.

Quine's suggestion that we try to make scientifically acceptable sense of meaning in terms of synonymy between expressions is declared inadequate by Dummett because:

'It is impossible to explain in general what is meant by "knowing (or learning) the meaning of a word" in terms only of the relation of synonymy.'

Thus far, we have only a vague characterisation of sense in terms of the two very general theses above. To understand what Dummett takes sense to be, we need to examine his conception of a theory of meaning within which the theory of sense plays a crucial role.

Dummett sets exceedingly high standards of adequacy for theories of meaning to satisfy. A theory of meaning must account for what a speaker

---

5 loc. cit. p.157
knows in understanding the expressions and sentences of his language in such a way as to show:

'not only how it (the language) does what it does but what it is that it does.'

In doing this, it must not employ any concepts which presuppose the notion of linguistic understanding or of being able to use a language at all. Only when we have constructed such a theory of meaning will we be able to command a clear view of what our highly complex linguistic ability consists in.

At Frege pp.413 -20ff. Dummett sets out his ideas on the form an acceptable theory of meaning should take. These ideas are later modified and refined in 'What is a Theory of Meaning?' and 'What is a Theory of Meaning (II)?', the conception of sense undergoing a transformation in the process:

A theory of meaning should make clear the roles played by the notions of truth and falsity in the language. Just as the rules of a game take for granted the significance of classifying certain states of play as 'winning' or 'losing' ones, a semantic theory for a language assumes the notions of truth and falsity as they feature in the systematic assignment it makes to the sentences of that language are likewise understood.

Now when the language in question is a formal one and our interest is in logical notions such as those of soundness and completeness this assumption does no harm. But when we are trying to understand the workings of a natural language, the mere assignment of truth-conditions to sentences is uninformative so long as we are uncertain as to how these truth-conditions

---

6 loc. cit. p.92.
7 Frege p.413.
are connected with the use to which their respective sentences are put in the actual practice of speaking the language.

Thus, a 'Russellian' language in which the predicate 'false' was applied to sentences with vacuous names and a 'Strawsonian' language in which such sentences were classified as 'neither true nor false' would only reflect a difference in the meanings assigned to sentences containing proper names (rather than a slight divergence in the meanings of the predicates translated as 'true' and 'false') if we could point to some difference in the use made of such sentences over and above the differing truth-values assigned to them: 'Until a connection is made between the truth-value of a sentence and the linguistic activities of asserting, questioning, etc., which can be accomplished by its utterance, we are in the dark as to what truth and falsity are, what is the difference between them, or what is the significance of ascribing them to sentences.'

Sentences only have a sense or express thoughts in virtue of the practices of using them to make assertions, ask questions, etc., according to Dummett. Dummett follows Frege in believing as against Wittgenstein that it is legitimate to separate the sense of an utterance from its force but claims that this is only warranted if it is possible to give a uniform description of the linguistic act that is effected by uttering an arbitrary sentence whose sense is presumed known with just that force.

Given that such a separation is justifiable, we can regard the theory of meaning as containing a pragmatic part - the theory of force - and a semantic part. The concept of truth will be relevant to both parts of the total theory: the semantic theory will provide a recursive truth-definition for the sentences of the language by specifying in its base clauses the semantic

8 loc. cit. p.415.
values of the primitive expressions of the language and rules for determining the semantic value of any complex expression from that of its parts; the theory of force will explicate the defined concept by locating it against the background of interests, activities, and attitudes that linguistic practice subserves.

So far, Dummett's view of the form a theory of meaning should take parallels the Davidsonian view sketched in §1.2, the only difference being that Dummett calls that part of the bipartite Davidsonian theory which specifies the contents of utterances, the theory of reference rather than the theory of sense. There is a reason for this differing terminology - Dummett believes that a theory of sense properly called is a theory of what the competent speaker of a language knows. So, alongside the inductive truth-definition, there will, on Dummett's picture, be a theory of (Dummettian) sense - call it a theory of senseD - which will supply an account of the cognitive aspects of an expression's use.

The theory of reference will act as input to the theory of senseD, specifying the reference of each expression and the theory of senseD will then say what a speaker grasps in grasping that reference. Dummett sums up his views in this way:

'The semantic part of the theory specifies the application of the notions of truth and falsity to sentences of the language (while the cognitive part shows how we recognise the application): the pragmatic part provides the point of so classifying sentences as true or false, by describing the use that can be made of any given sentence in terms of its truth-conditions.'

Now whilst 'sense can be taken only as a cognitive notion' Dummett is clear that:

\[9\] Frege pp.416-417.
\[10\] loc. cit. p.381.
'an account of sense does not aim to uncover an actual psychic mechanism.'

Rather, the account the theorist gives of sense:
'must be a theoretical model the test of which is its agreement with observable linguistic behaviour.'

The theory of sense is thus to be construed as a model of what the speaker knows in grasping the reference of an expression of any complexity.

Dummett allows at Frege p.381 that a theorist might conceive of certain notions employed in the model (in particular those of truth and falsity) as: 'theoretical notions, which cannot be correlated directly with linguistic behaviour, the model being judged correct or incorrect, according to its agreement with linguistic practice only as a whole.'

At 'What is a theory of meaning (II)?', however, Dummett is far less willing to make this concession. Here, the structure of and constraints upon acceptable theories of meaning are spelt out in finer detail. The competent speaker of a language obviously possesses the practical knowledge of how to speak the language,
'but this is no objection to its representation as propositional knowledge; mastery of a procedure, of a conventional practice, can always be so represented ... what we seek is a theoretical representation of a practical ability.'

The meaning theory will represent the practical ability as consisting in a grasp of a system of deductively connected propositions. As before, the semantic portion of the theory will centre around an inductive specification of the semantic values of each expression in the theory of reference. But now, surrounding this 'core' as a 'shell', the theory of sense:
'will lay down in what a speaker's knowledge of any part of the theory of reference is to be taken to consist, by correlating specific practical abilities of the speaker to certain propositions of the theory.'

So whereas before Dummett was content to regard a theory of sense with Frege as whatever it is the speaker knows in knowing the theory of

11 loc. cit.
12 loc. cit.
13 'What is a Theory of Meaning (II)?' pp.69–70.
14 loc. cit. p.74.
reference, he now requires the theory of sense to correlate known propositions of the theory of reference with practical linguistic abilities. This additional requirement is incurred, Dummett thinks, whenever we think of a theoretical model of a certain practical capacity as an item of implicit knowledge for the person whose capacity it models.

But isn't it highly implausible to attribute an implicit knowledge of a theory of meaning for his language to the average competent speaker? Since no one supposes a competent speaker of a language to be able to explicitly formulate a theory of meaning for his language, it is clear that we should ascribe at most an implicit knowledge of the theory to the speaker. But why ascribe such knowledge at all? It is the burden of Dummett's case in 'What is a Theory of Meaning (II)?' that the ascription to the speaker of a knowledge of the classical truth-conditions of a sentence is indeed vacuous.

The reason Dummett gives, to oversimplify, is that the clearest test for the possession of such tacit knowledge - viz. the speaker's recognising that the sentence is true when and only when its truth-condition is fulfilled - fails in the case of sentences which are not effectively decidable; and the only other models for what knowledge of a possibly recognition-transcendent truth might amount to also fail to generalise to the non-effective case.

Dummett goes on to argue that a theory of meaning which takes verifiability or falsifiability (rather than classical truth) as its central explanatory concept can pass the test above, since a speaker can be accredited an ability to recognise states of affairs which verify or falsify a sentence when they obtain.

Dummett's preferred model for a theory of meaning is a molecular one - one that pairs practical capacities with the theorems of the theory of
reference. As against this model, it might be claimed that there is simply no prospect of pairing anything less than the whole theory of reference with the practical ability to speak the language - that there is a problem in principle with compartmentalising this ability. So why can't the meaning theorist abstain from the task of pairing theorems with practical abilities, leaving this to the psycholinguist?

He can do so, Dummett argues, only if he is prepared to relinquish the claim that competent speakers of a language *implicitly know* the theorems of a theory of meaning for their language. For it is a general methodological principle, according to Dummett, that ascriptions of implicit knowledge must, if they are not to be vacuous, specify not only what the ascribee is held to know, but also how this knowledge is evidenced in his behaviour. Thus expressed, this methodological principle is surely correct - there is nothing intrinsically Anti-Realist about such a maxim, it is a sound means of controlling the ascription of content to putative tacit cognitive states. What plausible psychological theory could possibly violate it?

But do holistic theories really fail this condition? Doesn't Charles distinctively manifest his linguistic ability whenever Diana addresses him in English by being able to give primary redescriptions of her utterances? At this stage, we are in no position to decide this issue, at the very least we require a clarification of semantic holism and Dummett's precise objections to it.

Dummett claims that a non-classical molecular theory of meaning does warrant the attribution of implicit knowledge of a sentence's verification (or falsification) conditions to a competent speaker. His argument for this claim is this:

'It is not necessary that we should have any means of deciding the truth or falsity of the statement, only that we should be capable of recognising when its truth has been established. The advantage of

15 This is the task of S2.5
this conception is that the condition for a statement's being verified, unlike the condition for its truth under the assumption of bivalence, is one which we must be credited with the capacity for effectively recognising when it obtains; hence there is no difficulty in stating what an implicit knowledge of such a condition consists in - once again, it is directly displayed by our linguistic practice.

We have so far left the notion of implicit knowledge quite unanalysed and it might well transpire that it breaks down under analysis into a less clearly cognitive one - that of a skill, say (albeit a very intricate one). If so the assumption that the ordinary speaker had at some deep level internalised the propositions of an abstruse theoretical model of his semantic competence would appear dangerously misleading.

Dummett addresses this worry at 'What Do I Know When I Know a Language?'. He asks whether the practical knowledge attributed to the speaker explains the practical ability or whether the practical ability is all there is to the practical knowledge. Someone sympathetic to the latter will want to assimilate the case of language-learning to the learning of any other technique or skill - as Dummett puts it, for such a person "to know" in these cases means "to have learned".

Yet there is a contrast between knowing how to speak Spanish and knowing how to swim, Dummett contends - whilst it is conceivable that a person could just swim naturally without any previous training upon being thrown into the water, it is not conceivable that a person could just find themselves able to speak Spanish upon first arriving in a Spanish-speaking country. If a person were miraculously to find themselves uttering words which Spanish

16 'What is a Theory of Meaning (II)?' p.111, italics mine.
17 Even more so when the preferred model is an assertibility conditions theory rather than a truth-conditional theory, for such a theory might well construe 'knowledge of assertibility conditions' for a sentence p to be nothing more than knowing when one can assert p, which need involve no propositional knowledge (knowing that) at all.
18 Lecture to The Swedish academy, 1979.
19 loc. cit. p.2
speakers recognised as Spanish, this would not amount to a knowledge of Spanish. For as long as there was a doubt in that person’s mind as to whether they had interpreted a sentence at all (rather than whether they had interpreted it correctly) they could not be ascribed a knowledge of Spanish. Unlike skills such as swimming, there can be no gap between knowing what it is to speak Spanish and knowing how to do so.

Now clearly there are degrees of conscious involvement in the performance of different skills. So a fairer comparison, Dummett notes, would be between speaking a language and playing a game requiring reflective judgement such as chess. Dummett thinks that it would be possible for someone to learn to play chess by simply being corrected whenever he made an illegal move without ever being told the rules of the game. In that case, we should say that he had acquired a knowledge of the rules of chess, Dummett decides, rather than describe his new state as one in which he had merely acquired a technique. For:

"it would be unthinkable that, having learned to obey the rules of chess, he should not then be able and willing to acknowledge those rules as correct when they were put to him ... Someone who had learned the game in this way could properly be said to know the rules *implicitly* ... he does not merely follow the rules, without knowing what he is doing; he is *guided* by them." 21

Dummett gives us a tolerably clear guide to the application of ‘implicit knowledge’ - viz:

‘knowledge which shows itself partly by the manifestation of the practical ability, and partly by a readiness to acknowledge as correct a formulation of that which is known when it is presented.’ 22

Dummett then goes on to argue for the disputed thesis about knowledge of a theory of meaning:

‘A speaker’s mastery of his language consists, on this view, in his knowing a theory of meaning for it: it is this that confers on his utterances the senses they bear, and it is because two speakers take the language as governed by the same, or nearly the same theory of meaning that they can communicate with one another by means of that language.’ 23

21 loc. cit. p. 4
22 loc. cit. p. 3.
23 loc. cit. pp. 9-10.
I am not convinced by this argument of Dummett's. We can certainly agree that language-mastery implicates an agent in epistemic states otherwise unavailable to him, but the question is whether knowledge of the propositions of a meaning theory plays any role in the speaker's semantic competence and I do not see how Dummett's test for distinguishing between practical abilities is supposed to help us decide this issue. But this is a very vexed question.

Davidson, though not entirely pellucid on this issue, seems to disagree with Dummett here. Spanish speakers know the consequences of a truth theory for Spanish, that is, they know the propositions expressed by the T-sentences of a homophonic truth theory for Spanish, although this is a misleading way to put it, given the holistic character of their knowledge. A radical interpreter R can acquire this knowledge if he constructs from the evidence of Spanish speakers' behaviour an interpretative truth theory for their language. But his arcane way of acquiring this knowledge need bear (and surely won't bear) any psychological relation to the way native Spanish speakers themselves acquire their linguistic ability, even if both routes to mastery involve radical interpretation. Thus:

"Kurt utters the words "Es regnet" and ... we know that he has said that it is raining. ... we are able to ... interpret his words ... What could we know that would enable us to do this? How could we come to know it? The first of these questions is not the same as the question what we do know that enables us to interpret the words of others. For there may easily be something we could know and don't, knowledge of which would suffice for interpretation, while on the other hand it is not altogether obvious that there is anything we actually know which plays an essential role in interpretation. The second question, how we could come to have knowledge that would serve to yield interpretations, does not, of course, concern the actual history of language acquisition. It is thus a doubly hypothetical question: given a theory that would make interpretation possible, what evidence plausibly available to a potential interpreter would support the theory to a reasonable degree?"24

I think Davidson's way of seeing the relation between theories of the practical ability and the ability itself is by far the more attractive one, but

24 'Radical Interpretation', p.125.
we do not as yet know whether a Davidsonian (classical) truth theoretic model of a speaker’s semantic competence can ultimately be sustained. The theory of sense₀ plays a governing role in the total theory of meaning for Dummett. His favoured theories of meaning are molecular ones in which the theory of sense₀ specifies what a speaker’s knowledge of the T-sentences for each object-language sentence consists in by pairing these theorems with practical capacities to recognise the sentence’s truth-conditions as obtaining whenever they do. The theory of sense₀ thus constrains acceptable theories of reference to segment the practical ability to speak the language into such component abilities.

The theory of reference can be thought of as stating what a speaker knows, the theory of sense₀ as stating how this knowledge is manifested in behaviour. Hence Dummett seeks to explain the significance of language in terms of a speaker’s knowledge of his language. Now earlier I said that Dummett’s methodological principle governing the attribution of implicit knowledge ought to be acceptable to anyone. Yet Dummett insists on something much stronger than behavioural evidence for an ascription of tacit semantic knowledge, he insists that this knowledge should be exhaustively manifest in use, which demand converts an acceptable methodological principle into a quasi-behaviourist one.

What is Dummett’s justification for this stronger requirement? It is that meaning is essentially communicable and that for meaning to have the nature it has, it is necessary that what a speaker means by any expression be exhaustively recoverable just from the use that he makes of it. This is Dummett’s Manifestation Constraint and as we shall see it plays a pivotal
role in his argument against Realism. Dummett's clearest justification of this
demand occurs at 'The Philosophical Basis of Intuitionistic Logic' pp.216-217:
The meaning of a mathematical statement determines and is exhaustively determined by its use.
The meaning of such a statement cannot be, or contain as an ingredient, anything which is not
manifest in the use made of it, lying solely in the mind of the individual who apprehends that
meaning: if two individuals agree completely about the use to be made of the statement, then they
agree about its meaning. The reason is that the meaning of a statement consists solely in its role as
an instrument of communication between individuals, just as the powers of a chess-piece consist
solely in its role in the game according to the rules. An individual cannot communicate what he
cannot be observed to communicate: if one individual associated with a mathematical symbol or
formula some mental content, where the association did not lie in the use made of the symbol or
formula, then he could not convey that content by means of the symbol or formula, for his audience
would be unaware of the association and would have no means of becoming aware of it.'

The argument is meant to apply to all sorts of statements and not just
mathematical statements - indeed, Dummett's thesis is that the most
plausible ground for repudiating classical reasoning in mathematics for
intuitionistic reasoning is the wholly general version of the meaning-
theoretic argument above. I simply record it at this stage in order to
expound Dummett's position accurately.

Let us assume that Dummett is right about the central concept of a theory of
meaning having to be one that a speaker can recognise as obtaining
whenever it does. Let us also assume, as seems undeniable, that a speaker
comprehends the truth-conditions of a sentence by understanding the
component senses of sub-sentential parts. How, according to Dummett, does
the speaker derive his understanding of the sentence's truth-conditions from
an understanding of its parts? Dummett offers the Realist the following
model as an initially plausible one that might meet his strictures, deriving it
from Frege's writings on sense:
Since the central concept of the meaning theory has to be a recognisable one,
the sense of a sentence is to be thought of as a method for determining the
truth-value of that sentence. Moreover, if the speaker's understanding of
any sentence of his language is to be, as it must be, compositional, he must acquire an understanding of this method for determining the truth-value of the sentence from his understanding of the sentence's constituents.

Taking the simplest case of an atomic sentence first, a sentence of the form $Fa$, the sense of the singular term $a$ is comprised of some method of identifying an object as the referent of that term; the sense of the predicate $F$ as consisting in some means of determining whether a given object satisfies $F$. Having applied the two procedures, the speaker is in a position to determine the truth-value of the sentence.

Clearly this model will not generalise because there are many sentences for which we lack any method of determining their truth-value but which we understand perfectly clearly - the model only fits a decidable language. Yet our language is rife with undecidable sentences. On the Naive Realist model being mooted, we therefore appeal, in such cases, to the recognitional capacities of a superhuman being for whom undecidable statements would become decidable.
Dummett finds such an appeal incredible, as tantamount to an admission that we with our finite recognitional capacities, systematically misunderstand at least the undecidable sentences of our own language. But instead of dropping the requirement that a theory of meaning explicate the meanings of sentences of a language by means of a recognisable concept, he recommends that we drop the assumption of Bivalence - i.e. that every statement must possess one or the other of two truth-values. Having given up Bivalence, the model of understanding above now becomes defensible. Thus he writes:

'Thus, for the naive Realist, the connection between that which renders a statement true and our knowledge of its truth is an intimate one, just as it is for the Anti-Realist: from what it is like to know it to be true, we see just what it is for it to be true. Only they draw opposite conclusions. The Anti-Realist draws the conclusion that the statement cannot be true unless we know it to be true, at least indirectly, or unless we have the means to arrive at such knowledge, or at least unless there exists that which, if we were aware of it, would yield such knowledge. The naive Realist believes that the statement must be determinately true or false, regardless of whether we are able in the particular case, to perceive that which renders it true or false; but it is our capacity, in favourable circumstances, to perceive directly that which renders true or false other statements of the same type that constitutes our understanding of what it is for the given statement to be true or to be false.'

There is no dispute between the Realist and the Anti-Realist over the decidable fragment of a language according to Dummett. The dispute arises solely over sentences for which we lack any decision procedure. I shall refer to such sentences as 'non-effectively decidable sentences' ('non-ED' sentences for short). I will assume there are decidable empirical statements - statements whose truth-value can be definitely ascertained within a finite stretch of time.

What statements in a natural language could reasonably be thought of as non-ED? Wright says:

'multifarious types of sentences are in this situation: unrestricted spatial or temporal generalisations, many subjunctive conditionals, descriptions of the remote past, hypotheses about the mental life of others or of animals.'

Such statements, though presently undecidable might at some time be decided - it is just that we lack an effective way of bringing into being any deciding information. The Anti-Realist denies that there are any intelligible absolutely undecidable statements - statements about the properties of inaccessible ordinals, for example, are regarded as incoherent.

J.J.C. Smart's fantasy of a five-dimensional universe consisting of two causally isolated four-dimensional sub-universes would also be thought of as unintelligible - since we could in principle have no means of determining whether the truth-conditions of 'The Universe is five-dimensional' obtained, the Anti-Realist denies that we have a coherent grasp of its truth-conditions.

---

There is a principled basis for the rejection of such examples, at least if the Anti-Realism in question deploys an intuitionistic logic: this is simply that these cases lead to the assertion of a contradiction. For if we have grounds for asserting that it is logically impossible that $p$ should be verified, we thereby have grounds for asserting $\neg p$, and similarly if we have grounds for asserting that it is logically impossible that $\neg p$ should be verified, we thereby have grounds for asserting $\neg \neg p$; hence we have grounds for asserting a contradiction, to wit $\neg p \& \neg \neg p$ (which is intuitionistically equivalent to the denial of LEM). Thus the supposition that such cases represent coherent possibilities has to be rejected.

How do we then understand undecidable statements according to the Anti-Realist? By understanding their canonical verification conditions. We understand the meaning of 'Parmenides first walked on his first birthday' because we know what it would be to be in receipt of evidence conclusively attesting to this fact.\(^2\)

The Anti-Realist claims that our understanding of sentences is both compositional and molecular: we associate sentences with sets of assertibility conditions which confer an individual sense upon them and we recognise the meaning of sentences through familiarity with the meaning of their parts and structure. Thus the Anti-Realist need not insist implausibly with the positivists of yore that we understand a sentence only if we can conduct some test to verify or falsify it.

The Anti-Realist contends that truth cannot outrun our capacities to determine truth and recommends replacing truth with some non-

\(^2\) For example, we know what it would be to come across the hitherto undiscovered diaries of the famous monist's mother in which this fact was mentioned.
transcendent concept such as verification or falsification (or possibly both) as the central explanatory concept(s) of a theory of meaning. Is it necessary then for the Anti-Realist to allot any explanatory role to truth at all? Dummett has urged that it is: truth will no longer be the central concept in terms of which other semantic notions are to be explained, but it will play a fundamental role in the theory of meaning. The most important function it will fulfil is in giving an account of the valid inferences in the language. For there is no other way ultimately of characterising a valid inference than as one that preserves truth from premises to conclusion.3

Another possible reason for requiring truth to play an explanatory role in an Anti-Realist theory of meaning is this:

In a theory of meaning in which truth plays the central role, the content of any assertion is fully determined by the condition that the sentence uttered be true ... There is, however, no a priori reason why the truth condition of a complex sentence should depend only on the truth-conditions of its constituent sentences ... likewise, the general conception of a theory of meaning in terms of the conditions for the verification and falsification of a sentence carries no presumption that the meanings of the sentential operators will be explained in the comparatively simple way that they are in intuitionistic logic.4

While it does not follow from this observation alone that the notion of truth will be required to account for the meanings of the sentential operators, there are certain semantic intuitions which support this thought: my belief that a certain judge influenced the natural course of justice is more plausibly construed as a belief that the corresponding content sentence is true rather than warrantedly assertible; similarly the antecedents of conditionals seem to be hypotheses of truth rather than warranted assertibility, explaining the strong intuitive connection between truth and conditional statements that Dummett has so often emphasised.

3 Dummett makes this claim at 'What is a Theory of Meaning? (II)’p. 115 and argues for the primacy of semantic over syntactic characterisations of deducibility at 'The Justification of Deduction’ pp. 290–318 in T.O.E. and elsewhere. This is not to say that the best way to characterise the meanings of the logical constants may not be in proof-theoretic terms - cf. SS4.2, 4.3, 4.6.

4 'What is a Theory of Meaning? (II)’ pp. 112–113.
It will only follow from these considerations that:

'the doubt is exactly whether a systematic anti-realist theory of meaning is possible.\(^5\)

if we can demonstrate that the concept of truth is unavailable to the Anti-Realist. Contrapositively, an acceptable Anti-Realist meaning-theory ought, it seems, to include an account of truth.\(^6\)

Such a concept of truth would have to be one that speakers could in principle recognise as obtaining whenever it was held to obtain. It would not be necessary for truth to be effectively recognisable - 'recognisable "in principle" is a notion weaker than effective decidability, allowing for cases of recognition of truth being prevented by contingent misfortune or lack of ingenuity.\(^7\)

I now want to investigate an Anti-Realist proposal for accomodating the fact that most empirical statements are not conclusively verifiable (or falsifiable). Strawson has argued that statements about the mental life of others - 'Ludwig has a toothache', 'Cecile is thinking of Vienna' etc - together with statements about the past are simply not semantically explicable in terms of assertibility conditions or any type of Anti-Realist substitute for classical

\(^5\) Wright 'Truth Conditions and Criteria' p.235.

\(^6\) In fact, Dummett's case against Realism depends upon such an account, or so I shall argue in S4.2. Dummett is well aware of his obligation to provide such an account and acknowledges at 'What is a Theory of Meaning? (II)' p.116 that 'It is far from being a trivial matter how the notion of truth, within a theory of meaning in terms of verification, should be explained'.

\(^7\) One Anti-Realist unconvinced of the need for Anti-Realism to make out an acceptable notion of truth is Crispin Wright. He recommends a radical conventionalism to account for validity, according to which the validity of logical laws such as LEM is no longer explained in terms of the meanings of the logical constants which figure in them, instead, the laws we accept constitute the meanings of those very constants. The case for such a conventionalist view of meaning has been powerfully argued by Saul Kripke on behalf of the later Wittgenstein. I examine Kripke's arguments in chapter 5.
truth-conditions. This is because these statements relate to states of affairs that are inaccessible to the person uttering the sentence in question - he or she cannot directly experience Ludwig's toothache or Cecile's thoughts. Yet speakers have no difficulty at all in telling whether such assertions are supported by relevant evidence even though this evidence is held to be of necessity only indirect. How can this be if their semantic intuitions are not guided by a conception of these inaccessible states of affairs?

Wright has urged that the Anti-Realist must make use of Wittgenstein's notion of a criterion in explicating the semantic status of these claims; at the same time he has expressed scepticism that the notion as standardly interpreted is even coherent. The interpretation of 'criterion' that Wright discusses evinces three main features:

1. The Knowledge Feature (KF): Recognition that criteria for a non-ED statement p are satisfied confers knowledge that p upon an agent.
3. The Defeasibility Feature (DF): Any state of information in which the assertion of p is warranted by the satisfaction of its criteria can be expanded into a state of information in which the assertion of p is no longer justified if suitable evidence comes to light.

---

Wright's worry is that (DF) may undermine not only (KF) but also (MF). He provides a linguistic interpretation of criteria along the following lines: 10

Let \( D_1, D_2, \ldots, D_n \) represent decidable statements, \( P \) an undecidable statement.

Then we define a **Criteria Schema** as follows:-

If a speaker has verified each element of \( (D_1, D_2, \ldots, D_n) \) and has no information telling against \( P \) and no information which would explain why each element in \( (D_1, D_2, \ldots, D_n) \) is true unless \( P \) were true, then it is reasonable for him to believe \( P \). (CS)

Wright requires that \( (D_1, D_2, \ldots, D_n) \) be 'P-predicative' - that it contain no occurrences of \( P \) and the sentences \( D_1, D_2, \ldots, D_n \) do not require an understanding of \( P \) in order to be understood themselves. He then sets out to prove that for \( P \) a well understood undecidable sentence, a set \( (D_1, D_2, \ldots, D_n) \) can be found whose introduction with \( P \) into the Schema results in a non-contingent truth. If such a set can be found then there must be a set of conditions which can be explained independently of an understanding of \( P \) for which it is non-contingently true that their satisfaction defeasibly warrants an assertion of \( P \). Such conditions would be criteria for \( P \) evincing at least the features (MF) and (DF) above. His argument to that conclusion proceeds by reductio:

Suppose the opposite - i.e. that only a P-impredicative description of the assertibility conditions of \( P \) can be given. Then how do we acquire a grasp of the types of circumstances justifying the assertion of \( P \)? If only P-impredicative descriptions of these circumstances are possible, this can only be because either:

1. All P-predicative descriptions of \( P \)'s assertibility conditions only yield contingent instances of the Schema; or because

---

10 At 'Anti-Realist Semantics: The Role of Criteria' in *Idealism: Past and Present*, Royal Institute of Philosophy Lectures, 1978-1979 and also 'Second Thoughts'.

(2) It is impossible to give any P-predicative description of P's assertibility conditions at all. If (1) holds, then there can be no full characterisation of P's assertibility conditions since the demand that empirical grounds for the assertion of P be produced for any state of affairs putatively warranting it precludes any finite model being produced for how we recognise that P is assertible. We would, in other words, have to verify infinitely many descriptions of P's assertibility conditions before we could recognise any contingent P-predicative description of P's assertibility conditions as correct. Similarly, if (2) holds, there can be no full description of P's assertibility conditions either.

So, we must either reject (1) and (2), positing the existence of non-contingently true instances of the Schema - i.e., criteria - or else claim that no fully adequate description of P's assertibility conditions can be given - only ostensive characterisations are possible.

The assertibility conditions theorist is thus committed to the existence of criteria for any undecidable sentence whose assertibility conditions cannot be communicated ostensively - i.e., by demonstration and example.

But even for sentences whose meanings are given in this incomplete and ostensive manner, Wright argues, the theorist must at least be committed to an analogue of the notion of a criterion, which he calls a *criterion. A *criterion is a state of affairs that conventionally supports a given undecidable statement P, where the notion of conventional support comes to this:

A state of affairs X conventionally supports P is a speaker S is justified in asserting P on the grounds of X and we do not require S to know how to investigate whether X really does provide grounds for believing P in order to understand the assertion of P.
In contrast, \( \mathbb{N} \) would offer *contingent support* for \( P \) if we did require \( S \) to know how to empirically ascertain whether \( \mathbb{N} \) really did provide grounds for asserting \( P \).

Wright again proceeds by reductio in order to prove that the assertibility theorist is committed to *criteria for ostensively explained assertibility conditions for non-ED \( P \):

If the theorist rejects both criteria and *criteria for such sentences, he can only explain the meaning of \( P \) by \( P \)-predicative descriptions that are only contingently true if true or by ostensively explained assertibility conditions that offer only contingent support for \( P \). In either case however, speakers who understand \( P \) will know how to test the characterisation empirically and thus these characterisations might prove incorrect.

Wright concludes that if a full understanding of the assertibility conditions of \( P \) is to be communicable at all, it must be achieved by recourse to either criteria or *criteria.

I have outlined Wright's argument in some detail because I think it is very important for the evaluation of the feasibility of a global Anti-Realist semantics. I think Wright's argument is sound. We wish to know not just that a speaker asserted \( P \) when a certain condition in fact held, but that he asserted it *because* the condition held.\(^{11}\) Failing an apriori pairing of \( P \) with canonical assertibility conditions, it is unclear what a speaker might be saying in asserting that \( P \) on an assertibility-based semantics. Moreover, for empirical \( P \), these canonical conditions would seem to be essentially

\(^{11}\) A similar point was made in S1.3 apropos Davidson's use of material biconditionals in the Tarskian schema (\( T \)) \( s \) is true iff \( p \).
defeasible - any conceivable such conditions for 'Ludwig has a toothache' such as Ludwig's groaning and holding his jaw would only verify that Ludwig really has a toothache on the assumption that subsequent evidence that he was only shamming will not be forthcoming.

Essentially defeasible canonical assertibility conditions just are conditions satisfying (MF) and (DF) and it would be a very implausible version of Anti-Realism which chose to deny that we can ever know whether another person is in pain or whether a certain past event occurred or not; hence (KF) ought to be preserved. The conclusion that the Anti-Realist is committed to the existence of criteria, then, seem highly plausible.

Yet if this is so, the prospects for a credible Anti-Realist theory of meaning look bleak. For the notion of a criterion appears to be (and both Wright and McDowell have persuasively argued that it is) incoherent. Surely, if Norman does not know how to test whether Ludwig's criteria jaw-holding really does result from toothache, Norman does not really understand the meaning of 'Ludwig has a toothache'?

The criteria-theorist might respond that Norman need only recognise that the criteria are defeated whenever they are; he need not have any idea how to establish that the 'conventionally' determined criteria really do justify the assertion of the corresponding undecidable P.12 But this counter fails of plausibility so long as we think of the relation between undecidable P and its criteria to be one of explanation - the intuitive response that assertions about empirical states of affairs, irrespective of whether they are decidable or undecidable, can only be warranted by empirical evidence, remains unchallenged: the whole notion of a special type of evidence which non-

---

12 Wright tentatively endorses this line of response toward the end of 'Second Thoughts'.

inductively justifies the assertion of statements about the past or about the mental states of others as a matter of 'grammar' seems a complete sham.13 Still, there might be a Realist assumption unfairly being appealed to here. The criteria-theorist claims that my assertion that Ludwig has a toothache is justified upon recognition that criteria for this claim are satisfied.14 We must not, he holds, construe my knowledge-claim that Ludwig has a toothache as knowledge that 'Ludwig has a toothache' is true, even in optimal circumstances since this would be to assimilate criteria to the very truth-conditions they are invoked to replace.

But then if the criteria for 'Ludwig has a toothache' do not determine whether the knowledge claim that he has is true or not, what does? It would seem that the circumstance of Ludwig's really having a toothache (i.e. the corresponding knowledge claim being true) lies outside the speaker's epistemic grasp on the criteria theorist's showing; so how can the theorist justify his reliance upon the satisfaction of criteria as an indication of its

13 It is no reply to this to claim that the convention which it is senseless to question is that we take the sincere avowals of others as the ultimate arbiters of their mental states, since these avowals are 'P-impredicative'. Moreover, how are we to decide that an avowal is or isn't sincere on this behaviourist view?

14 Compare Wright's response to McDowell's suggestion that the circumstance of another's being in pain was sometimes available to one's awareness directly, unmediated by inferences from behavioural proxies: 'But that no inferences, via 'proxies' or whatever, should be involved is quite consistent with what is actually perceived being not that someone is in pain, tout court, but that criteria... that he is in pain are satisfied. Criteria are not proxies, and they do not form the bases of inferences, correctly so described. But in contrast with truth-conditions, a claim made on the basis of satisfaction of criteria can subsequently be jettisoned consistently with retention of the belief that criteria were indeed satisfied.' at 'Realism, Truth-value links, Other Minds and the Past' p. 123 in Ratio 22, 1980, pp. 112-132.
presence? Better to admit that we just cannot know what others are experiencing or even whether they have experiences at all.\textsuperscript{15}

But perhaps there is an illicit Realism tacitly in operation here. For it still seems as if we are after the truth-conditions of knowledge claims whereas we should simply describe the criteria for making them and the types of situations which would subsequently defeat them.

As Wright clearly shows, this response does not extricate the theorist from his difficulties. I will surely be taken to have misled my audience if in asserting that $P$ in a criterial state of information (CSI), I do not rule out the subsequent emergence of a defeating state of information (DSI). For if such a DSI does emerge and I regard it with equanimity, as disappointing no expectation that I harboured at the time of issuing my claim to know that $P$, then I must admit that in CSI it seemed to me that all available evidence I had in support of $P$ was consistent with things not being as I claimed they were; and this is just to admit that I did not know that $P$, despite being accredited such knowledge by the theorist.

Lowering the demand from knowledge to reasonable belief consistently with the abnegation of truth is no more successful - the problem is that CSI and DSI are distinct states of affairs and neither the possibility of a DSI arising later nor the frequency with which such an event might be expected to occur are at all predictable from my epistemic standpoint at CSI. It is thus quite possible for a certain criterion $c$ for non-$EDP$ to be defeated so often by circumstances that we should rationally cease to regard $c$ as a ground for

\textsuperscript{15} One radical response to this criticism that Wright envisages is for the theorist to give up the requirement that knowledge entails truth and construe criterially justified knowledge claims as covertly indexical. Thus "I know that $P$" is to be construed as the claim that one's present state of information includes a recognition of the satisfaction of criteria for $P$ and of no defeating evidence. The obvious difficulty here is that the theorist is thereby forced to admit that knowledge can be lost and to thus effect a complete break both with traditional conceptions of knowledge and with anything resembling intuitionistic semantic models which assume that knowledge is cumulative.
asserting $P$ at all. In that event, however, the claim that $c$ provided 'necessarily good evidence' for $P$ would just be false and the a priori pairing of $c$ with $P$ as the method of fixing $P$'s meaning would surely have to be revoked.

Thus, as Wright contends, the essential defeasibility of criteria undermines not only the Anti-Realist's claims to knowledge or reasonable belief about the truth of statements concerning the past or other minds that he might venture on the basis of satisfaction of criteria in a CSI, but also the power of criteria to specify the meanings of non-ED statements.
Our task is now to outline Dummett's case against Realism. Realism about a class of sentences is the view that the meanings of those sentences can be explicated by a classical two-valued semantics together with a truth-conditional theory of meaning based upon that semantics. Opposed to that thesis stand a variety of Anti-Realisms - the most radical denying that a truth-conditional account of their meanings can be given at all; the least radical denying that reference plays any role in an explanation of their truth or falsity. The denial that Bivalence is generally warranted lies somewhere between these two extremes. As this latter version is the one Dummett has most explicitly opposed to Realism, I will discuss it, but I think his arguments can easily be generalised to the other variants.

Anti-Realism is first and foremost the denial that classical truth and falsity (truth subject to Bivalence) can be the central explanatory concepts in the theory of meaning; it carries no positive commitments at all - in particular, it ought not to be identified with the claim that some other non-verification-transcendent notion of truth such as intuitionistic truth should play that role.¹

We now only need to review Dummett's conception of the structure of a theory of meaning before presenting his case against Realism:

¹ Recall that for Dummett truth will still play a crucial role in any plausible Anti-Realist alternative to Realist semantic theory but that it will no longer be the primitive explanatory concept in that theory, being defined instead in terms of some alternative notion such as verification or falsification which does not transcend our recognitional capacities.
(1) The 'core' of a theory of meaning is the theory of reference which consists of a finitely axiomatised Tarskian truth theory assigning references to the semantic primitives of the language under study and recursively specifying the truth-conditions of its sentences. This core provides a theoretical representation of what the competent speaker of the language knows.

(2) Surrounding the core as a shell is the theory of sense. This provides a model of what the speaker's knowledge of any part of the theory of reference consists in by correlating certain practical linguistic abilities with known propositions of the theory of reference. An atomic theory of meaning will correlate these abilities with the axioms of the truth theory; a molecular theory will correlate these abilities with the theorems of that theory; a holistic theory will attempt no partitioning of the theory and will take the whole theory to provide a model of the undifferentiated practical ability. This shell thus provides a theoretical representation of the manner in which the speaker's semantic knowledge is made manifest in his linguistic practice.

(3) Finally, the theory of force will provide an account of the conventional significance an utterance of a sentence in a given mood has - it will demonstrate the kind of linguistic act typically performed by an utterance of a sentence in a designated mood.

It is the theory of sense that is pivotal in Dummett's case against Realism for it will rule out as unacceptable any meaning theory which does not permit a mapping of implicitly known propositions of the theory onto recognitional abilities manifesting that knowledge. We are now in a position to set out Dummett's argument against Realism.2 Dummett makes three assumptions that serve to constrain acceptable theories of meaning:

2 The version I present here is Dummett's Manifestation Argument as it is most fully expounded in 'What is a Theory of Meaning? (II)'. I take this argument against Realism to be more important than his Acquisition Argument which seeks to discredit Realism through considerations of what is involved in acquiring a language. My justification for (temporarily) ignoring the latter is that, on Anti-Realist premises, even if we could acquire a knowledge of classical truth-conditions, the question would still arise as to how we could be certain that we managed to communicate this knowledge coherently to our fellow speakers. I return to the Acquisition Argument in the context of discussing Dummett's argument for revision of logical practices at §4.2 and §4.5.
\( A_1 \) The meaning of any statement (or expression) is exhaustively determined by and exhaustively determines the use of that statement (expression). The reason for this is that language is essentially a vehicle for communication. To suppose that meaning could somehow transcend use is to make meaning ineffable, according to Dummett.

\( A_2 \) In the light of \( A_1 \), we should characterise semantic competence as the practical ability to use the expressions of a language in appropriate circumstances. Though practical the ability is nonetheless cognitive. But the knowledge required to be able to speak a language aright cannot in general be explicit verbalisable knowledge on pain of regress or circularity. It must therefore be implicit knowledge.

\( A_3 \) Attributions of implicit knowledge are only justifiable if we can say what the manifestation of that knowledge consists in - that is, what observable difference the possession of such knowledge makes to the behaviour of the person it is ascribed to.

As the case against Realism is easier to grasp with those background assumptions in mind, I shall summarise these as:

\( R_1 \) Meaning is exhaustively determined by and determines use.

\( R_2 \) Semantic knowledge is implicit knowledge.

\( R_3 \) Implicit knowledge must be manifestable in behaviour.

Now consider the Realist's suggestion that the meaning of a sentence is given by its truth-conditions. By the constraints above, an acceptable truth-conditional theory of meaning is one in which a speaker's implicit knowledge of the truth-conditions of the sentences of his language determines the use he makes of those sentences and is manifestable in his linguistic behaviour. The interpreter must then defend his attribution to a speaker of a knowledge of truth-conditions which is largely implicit.
(1) For effectively decidable sentences, the ascription to a speaker of implicit knowledge of their truth-conditions can be justified if the speaker can recognise whether or not those conditions obtain within a finite time.

(2) For non-EO sentences however, this type of test clearly cannot apply since the truth-conditions may obtain or fail to obtain without the speaker’s recognising this fact. But just because such truth-conditions are recognition-transcendent, no practical ability on the part of the speaker can be involved which manifests a grasp of them. So, the attribution of implicit knowledge of these conditions is wholly vacuous.

(3) The only type of knowledge that a speaker can have of the truth-conditions of non-EO sentences is thus explicit, verbalisable knowledge which is useless here.³

(4) But we clearly do understand the various types of non-EO statements speakers utter on occasions, since we know the types of circumstances in which their assertion is warranted.

(5) Realism, however, makes the attribution of such knowledge to speakers indefensible. Hence, as speakers undoubtedly do possess such knowledge, Realism must be false.

Supposing we grant, pro tempore, (A₁), (A₂) and (A₃). What should we make of premises (1) to (4)? Premise (3) might seem dubious - it is not by any means obvious that explicit knowledge of non-EO sentences is useless. Premise (2) also looks questionable. Why should the practical abilities that manifest a grasp of truth-conditions be concerned solely with the recognition of the fulfilment of those truth-conditions? Why isn’t a language-user’s ability to interpret another speaker’s utterances, or more precisely, in the terminology of §1.3, to provide primary redescriptions of another speaker’s utterances, practical ability enough, irrespective of whether the speaker’s statements have verification-transcendent truth-conditions? Premise (2) seems to assume without justification that a sufficient condition for the ascription of implicit knowledge in (1) is a necessary condition as well and Realists naturally deny that this is so.⁴ We plainly need to look more closely at the support Dummett provides for these premises.

³ 'What is a Theory of Meaning? (II)’ p.82
Dummett thinks that there are just two models for a speaker's knowledge of the truth-conditions of a sentence:-

(A) Explicit verbalisable knowledge
(B) The capacity to observe whether the truth-condition holds or not.

The application of the concept of truth to sentences is controlled by two guiding principles, Dummett contends in 'What is a theory of meaning (II)?:

(a) Principle C - if a sentence is true there must be something in virtue of which it is true; and:
(b) Principle K - if a sentence is true, it must be possible in principle to know that it is.

Dummett connects the two principles thus: if it were impossible to know the truth of a given sentence, then there could be nothing which made that sentence true.

Dummett clearly does not believe that these principles tell us the whole truth about truth, however I doubt that they tell us nothing but the truth either. What does Principle C mean? Dummett surely doesn't intend it as a statement of the Correspondence Theory of Truth, since he believes that to be false. His idea seems to be that the attribution of 'true' to a statement must be justifiable on some ground or other, rather than there being some

5 loc. cit. p.97.
7 See 'Can Truth be defined?' in Frege.
state of affairs in the world whose obtaining 'makes' a given sentence true.\footnote{Cf. also Davidson's remark at 'In Defence of Convention T' loc. cit. p. 70: 'The original question is not confused, only vague. It is: what is it for a sentence (or utterance or statement) to be true? Confusion threatens when this question is reformulated as, what makes a sentence true? The real trouble comes when this in turn is taken to suggest that truth must be explained in terms of a relation between a sentence as a whole and some entity, perhaps a fact, or state of affairs. Convention T shows how to ask the original question without inviting these subsequent formulations. The form of T-sentences already hints that a theory can characterise the property of truth without having to find entities to which sentences that have the property differentially correspond.'}

I find it hard to grasp this talk of 'making true' under any of its interpretations. For the Fregean reasons discussed in §1.2, I find the correspondence theorist's 'explanation' as to why p is true in terms of its corresponding to a state of affairs \(W\) unenlightening. I am happy to admit such relations of correspondence, but baulk at the suggestion that they explain what the truth of a true statement consists in. Dummett, by way of contrast, seems to baulk at the correspondences whilst agreeing that we need some truth-maker or other. Why does he think we need this? I suspect that it is because he believes that:

Unless there were some explanation as to what the truth of a true statement consists in, our practice of applying the concept of truth must really be incoherent.

We should distinguish an:

(i) Explanation as to why a statement \(p\) has a certain semantic property, from:

(ii) Explanation as to what it is for a statement to have that property, what having that property consists in.

Explanations of form (i) allow the semantic property in question to be irreducible, those of (ii) do not since to admit that a property is irreducible is precisely to admit that we have no explanation of what having that property consists in. The difficulty with Dummett's Principle C and with all talk of something's making a statement true is that it blurs the distinction between (i) and (ii) - if Principle C means that for any \(p\) which we regard as true we
must be able to explain why \( p \) has the property of truth, then this varies
from the truistic to the false depending on what we count as an acceptable
explanation: if an acceptable explanation is just one wherein we adduce some
reason for believing \( p \) (perhaps \( p \) itself) then it is truistic; if it is one in which
we are brought to a position wherein we see why it has to be the case that \( p \)
(as in a deductive-nomological explanation of empirical \( p \)), then it is false –
there is no reason at all to believe that we must be capable in principle of
explaining why the fundamental laws of physics hold, for example.

If, on the other hand, Dummett intends, as I think he does, (ii) rather than (i)
by Principle C, the principle ought to be rejected on the Fregean grounds that
there can be no reduction or deep explanation of what the property of truth
consists in. Davidson puts the point admirably:

‘Nothing, however, no thing, makes sentences or theories true: not experience, not surface
irritations, not the world, can make a sentence true.’

Principle K might cause the Realist some disquiet since it asserts that if a
sentence is true at all it must be possible to know that it is. But Dummett
does not mean to restrict the notion of possibility to human possibility at this
stage. The Realist can agree to the logical equivalence of ‘\( p \)’ and ‘an
omniscient agent would know that \( p \)’.

We can sum up the import of Principles C and K in a way which preserves
the ambiguity of C as follows:

In order for a sentence to be true, it is necessary that there be some
knowable ground for its truth.

Dummett calls a sentence \( P \) ‘barely true’ if it is true and there is no reductive
class of sentences for any class of sentences containing \( P \). A sentence which

---

cannot be barely true has some reductive class of sentences \( R \) such that an utterance of that sentence can only be true if all the sentences in some subset of \( R \) are true.

Dummett maintains that there is no problem in ascribing to a speaker a knowledge of a non-barely true sentence - either the theory of reference will provide a 'non-trivial' \( T \)-theorem for it, articulating the way in which its truth depends on the prior truth of some subset of sentences in its reductive class or, because of some problem in translating the sentence into the metalanguage, the \( T \)-theorem will be of a 'trivial' form (e.g. 'snow is white' is true iff snow is white) and it will then be the task of the theory of sense to describe the relationship between the sentence and its reductive class.\(^{10}\)

It is barely true sentences that cause the problem. For with a barely true sentence the corresponding \( T \)-sentence must be trivial and so the whole burden of explaining what it is to grasp the meaning of such a sentence falls on the theory of sense. However, the only plausible model for a speaker’s understanding of a barely true sentence, Dummett contends, is an observation report.\(^{11}\) But however generously we extend the notion of a sentence whose truth can be settled by observation, we cannot include within its compass sentences quantifying over infinite or unsurveyable domains, counterfactual conditionals, sentences referring to inaccessible regions of space or time or any other non-ED sentence.

Hence, if the Realist permits certain non-ED statements to be barely true, he will be bereft of a model which could justify the attribution of implicit knowledge of their truth-conditions to a speaker.

\(^{10}\) "What is a Theory of Meaning? (II)" p.95.
\(^{11}\) loc. cit.
Now Dummett does not say why an observation report is the only model available to the Realist here, but it is clear that wherever the truth of P is dependent on the prior truth of \( D_1, D_2, \ldots, D_n \) in some R-class, the ascription of a grasp of P's truth-conditions to a speaker is defensible provided we are already convinced that he understands the truth-conditions of those sentences in the R-class. But this clearly cannot proceed indefinitely - at some point the Realist must allow that a given statement is barely true and that its truth-conditions cannot be informatively stated. It is at that point, according to Dummett, that the demand for behavioural manifestation of this piece of semantic knowledge becomes difficult to satisfy. Observational reports are the only subclass of homophonic statements of P's truth-conditions which escape the charge of vacuity. So, Dummett concludes, there are only two sorts of acceptable manifestations of a knowledge of truth-conditions - (i) explicit, verbalisable knowledge; and (ii) The capacity to observe whether the truth-conditions hold or not. But neither of these are appropriate models for a putative grasp of the truth-conditions of non-ED sentences.

Now the whole force of Dummett's argument depends upon the vacuity of putative models for an understanding of Realist truth-conditions. A Realist might respond to the argument either by questioning Dummett's claim that explicit verbalisable knowledge is inadequate as a model of this understanding or by questioning Dummett's restriction of legitimate manifestations of such knowledge to just the verbal and the observational. Normally we do credit speakers with an understanding of a statement if they evince a sensitivity to the types of evidence that would confirm it and to the network of implications that the statement sustains. So why doesn't
Dummett allow these abilities on the part of a speaker to manifest his grasp of a sentence's truth-conditions?

The short answer to this is that he does. But what makes the attribution of understanding defensible in these instances is, he would argue, our tacit conviction that the speaker would recognise the truth-conditions as obtaining whenever they did (or that the speaker can relate these truth-conditions to other recognisable truth-conditions of sentences in some suitable reductive class). Dummett plainly regards a speaker's evidential evaluations as derivative upon his grasp of the sentence's meaning:

'... the principles which govern what counts as evidence for the truth of a sentence must be systematically derivable from its meaning.'

So, while Dummett could agree that evidential and inferential abilities on the part of the speaker do give us reason to ascribe a grasp of a sentence's meaning to a speaker, these abilities are derivative upon a verbal or observational grasp of truth-conditions - they are part of what a theory of meaning has to explain.

Dummett's question, then, is how grasp of recognition-transcendent truth-conditions can determine that certain evidence is good evidence for the truth of a given statement and how mere reasoning in accord with classical logical principles can answer his challenge to show that the ascription of implicit

---

12 'What is a Theory of Meaning? (II)' p. 132. Dummett would thus rule out as unacceptable not only a 'conceptual role semantics' in which a speaker's understanding of all sentences was determined by his grasp of their evidential and inferential connections rather than their truth-conditions, but also such a semantics for the non-ED part of a language.

13 Cf p. 132, op. cit.: 'What we are concerned with is what determines that something is evidence of a certain strength for the truth of a given sentence, not with whether the existence of evidence of a particular strength is sufficient reason for accepting the sentence as true.' This is another point at which Dummett and Wright disagree. Wright has the Anti-Realist respond by asking why the Realist should seek an explanation of our evidential or inferential abilities, claiming that we just do have the practical ability to discriminate states of information justifying the assertion of p from states of information not justifying its assertion. Because he believes that the ascription of implicit semantic knowledge is of genuinely explanatory value, Dummett demurs from this - these abilities precisely are ones that require explanation though not in the terms envisaged by the Realist.
knowledge of recognition-transcendent truth-conditions to a speaker is not vacuous.

Summary
Let us just review the course of Dummett's manifestation argument against Realism:

(1) Dummett lays down three constraints on any acceptable theory of meaning:

(A_1) **Meaning is exhaustively determined by and determines use.**

(A_2) **Semantic knowledge is implicit knowledge.**

(A_3) **Implicit knowledge must be manifestable in behaviour.**

(2) In accord with these constraints, an acceptable Realist theory of meaning must show how a speaker's implicit knowledge of the truth-conditions of a sentence both determines the use he makes of that sentence and is manifest in his behaviour.
(3) The argument then proceeded:

(P1) Ascription of an implicit knowledge of an ED sentence's truth-conditions is unproblematic because the speaker can recognise these truth-conditions as obtaining whenever they do.

(P2) But for non-ED sentences, such attributions are vacuous just because the speaker cannot recognise their truth-conditions as obtaining whenever they do and such recognitional ability is the only practical ability which would demonstrate such a grasp.

(P3) Ascribing explicit knowledge is equally useless.

(P4) But we clearly do understand the assertibility conditions of non-ED sentences.

(C) Realism must therefore be false.

(4) We saw that (P2) and (P3) looked questionable. In particular, (P2) seemed to treat what was acknowledged as a sufficient condition for the ascription of implicit knowledge as a necessary condition also.

(5) We then noted that Dummett's Principles C and K afforded some reason for identifying understanding with a recognitional ability - for according to those constraints, if a sentence is true at all there must be some knowable ground for its truth. The problem was that Principle C was ambiguous and under its non-truistic readings, seemed either dubious or false.

(6) The request for such a knowable ground led to the narrowing of acceptable models for understanding to just two - a verbal model for sentences whose truth depended on the prior truth of sentences in some reductive class and an observational model for 'barely true' sentences.

(7) It was then seen to be a moot question whether these models were exhaustive of our conception of semantic knowledge, the Realist objecting that evidential and inferential abilities should also be included.

Of the assumptions, (A1) and (A2) seem to be ones that the Realist might well take issue with. (A3) is the entirely reasonable requirement that ascriptions of implicit knowledge (whether of meaning or of anything else) be borne out by behaviour. It is the quasi-behaviourist demand of (A1) that such knowledge be exhaustively manifest in use that the Realist could and I believe should contest. However, (A1) is the conclusion of Wittgenstein's private language argument (at least as Dummett interprets it).
A satisfactory rejoinder to Dummett must therefore either show Wittgenstein's argument to be unsound or show why Dummett's interpretation of it is mistaken or else show that the argument does not have the Anti-Realist implications Dummett takes it to have.
Dummett's Manifestation Argument rules out as unacceptable any conception of meaning that cannot be manifested in use. According to this argument, we can justifiably ascribe implicit knowledge of a sentence's truth-condition to a speaker either when the sentence cannot be barely true and we are confident that the speaker understands the relevant subset of sentences in the reductive class for it or when the sentence can be barely true and the speaker can observe it to be true.

Any Realist position that denies that non-ED statements can be barely true will comprise a defensible response to the Manifestation Argument provided that it can produce for each type of non-ED sentence a reductive class of statements whose truth-conditions a speaker can unproblematically be credited with grasping.¹ Holistic theories of meaning do in fact deny that there can be barely true non-ED statements since they deny that any statement can be barely true. Dummett is thus concerned to establish that such theories of meaning are untenable.

I shall commence by outlining Quine's version of holism and then recall Davidson's version. After that we shall be in a position to set out Dummett's objections to these views.

¹ This condition would not be satisfied in the case where the non-ED statement P has a reductive class with an infinite number of decidable statements as its members. For while P would then not be barely true and while every member of the R-class is decidable, statements quantifying over such an infinite totality would then themselves be non-ED (of the same type as statements quantifying over an infinite domain of natural numbers).
We have already seen (in chapter 1) that the evidence available to Davidson's radical interpreter is the overt behaviour evidence of the sentences a native holds and prefers true. Because he thinks that there are no psychophysical laws, Davidson believes that neurophysiological facts or facts about functional organisation are of no use to his interpreter. Quine is even more insistent that Dummett's manifestation constraint be met: 'Language is a social art which we all acquire on the evidence solely of other people's overt behaviour under publicly recognisable circumstances.'

Whatever the shortcomings of Davidson's and Quine's versions of holism, then, violation of the Manifestation Constraint does not appear to be amongst them.

What makes Quine's theory of meaning holistic is its adherence to Duhem's Thesis - theoretical sentences only have their meaning and evidence when taken together; they face the tribunal of experience collectively. Duhem's thesis is a claim about the theoretical sentences of a scientific theory and their relation to experience; Quine extends the thesis to the sentences of a natural language which are in the main theoretical by his reckoning.

To understand Quine's holism, we need to see why he assimilates languages to theories. At 'The Nature of Natural Knowledge' p.68, Quine asserts: 'Our only source of information about the external world is through the impact of light rays and molecules on our sensory surfaces.'

Quine believes a theory of meaning must observe the crucial consequence he draws from this fact: viz. that science is the sole arbiter of truth and meaning. Indeed, there are two cardinal tenets of empiricism that Quine finds 'unassailable':

2 'Ontological Relativity' p.26 in Ontological Relativity and other essays, Columbia University Press, New York, 1969. Cf 'Semantics is vitiated by a pernicious mentalism as long as we regard men's semantics as somehow determinate in his mind beyond what might be implicit in his dispositions to overt behaviour.' (p.27) and '...there are no meanings nor distinctions of meaning, beyond what are implicit in people's dispositions to overt behaviour.' (pp. 28-29).
It is through their connection with sensory evidence that the sentences both of natural language and of science ultimately derive whatever objectivity and factuality they possess. Now the acquisition of a language and the acquisition of a theory of the world are, Quine contends, inseparable - 'we're working up our science from infancy on.' Certain linguistic expressions - occasion sentences - are directly linked to non-linguistic reality via simple conditioning; a subclass of occasion sentences - observation sentences - are expressions we've learnt to associate holophrastically with publicly observable concurrent circumstances. It is: 'through these expressions that language and science imbibe their empirical content.'

The language learner has successfully mastered the use of an observation sentence when he knows when to expect a veteran speaker to approve his utterance of it. The success of such conditioning depends on the existence of shared standards of similarity - a fact attested to by the common pattern of expectations evinced by veteran and novice. Quine admits that 'direct conditioning or simple induction does not suffice for the acquisition of language generally.'

The use of more complex grammatical constructions such as those involving tenses or modality, the use of metaphors or theoretical terms all elude instruction by this method. Quine simply claims for such cases that: 'the superstructure is cantilevered outward from that foundation by imitation and analogy, by trial and error.'

---

5 The Roots of Reference, La Salle: Open Court, 1974, p. 138.
7 loc. cit. p. 158.
8 loc. cit.
Sensory stimuli thus form the common base from which our language and our science develops. But even where an observation sentence is firmly conditioned to sensory stimuli, Quine maintains, there is no way of telling whether the language learner has acquired this sentence by direct conditioning to the relevant stimuli or by indirect conditioning through the transitivity of the conditioning of old connections of sentences with sentences. Hence the non-theoretical empirical content of the speaker's language cannot be isolated.

If we now go along with Quine in regarding 'a man's theory on a given subject ... as the class of all the sentences ... that he believes true', it becomes clear that no distinction can usefully be drawn between those sentences the speaker holds true because of their logical and semantic relations to other sentences (indirect conditioning) from those he holds true because they are confirmed by specific types of sensory stimuli (direct conditioning).

In this manner, Quine is led to a holistic theory of meaning for the sentences of a language generally and not just for its scientific subpart - these have their evidence and meaning only when taken together just as the theoretical statements of physics do and for just the same reason: there can be no isolating the empirical content of either. As Quine himself puts it:

'the evidence relation and the semantic relation of observation to theory are co-extensive'.

Dummett thinks that this holistic theory precludes the possibility of a systematic theory of meaning. He finds the Quinean metaphor of language as an articulated web of inferentially connected sentences with experience impinging only upon its periphery 'useful and licit' and he wholeheartedly

---


10 The Roots of Reference p.38.
supports the verificationism Quine subscribes to. His worry is that Quine's holism ultimately subverts both of these insights.

Dummett's objection is to Quine's generalisation of Duhem's Thesis - the suggestion that some sentences of our language have no independently ascertainable assertibility conditions, 'no fund of experiential implications they can call their own.'

Semantic holism is often identified with a network model of meaning whereby a sentence's meaning is identified with its location within the inferential network. Dummett has no objection to the network per se – indeed, he insists both that natural languages do form such structures and that these structures are or approximate to partial orderings. He just rejects the Quinean account of how this network relates to experience and as a result of that cannot agree that the meaning of a sentence is constituted by its inferential connections. What transforms the network metaphor into a holistic theory are the related theses that no experience compels the rejection of any sentence and that no sentence is immune from revision.
Dummett draws several unwelcome consequences from these theses -
(i) That they destroy the internal structure of the network, obliterating the
distinction between periphery and interior and generally making it unclear what
constitutes the totality of our linguistic practices.\(^{14}\)

(ii) That there is something arbitrary and unwarranted in the representation of a
speaker's linguistic dispositions by a constant theory of meaning - there is no
way of discriminating between one sentence being held true according to the
reasons S had for accepting it and any other sentence's being accepted as true: no
possible behavioural manifestation could reveal this.\(^{15}\)

(iii) That a holistic theory of meaning is tantamount to a denial of the possibility
of a theory of meaning for a language since 'no model can be given of the
individual contents of sentences.'\(^{16}\)

Now (iii) seems question-begging since a holistic theory of meaning such as
Davidson's precisely does deny that sentences have 'individual contents' yet
purports nonetheless to provide a theoretical representation of a speaker's
linguistic capacity.

Objections (i) and (ii) are more serious. Dummett argues that Quine's theory
has these consequences at Frege p.596:
'Quine's thesis involves that the principles governing deductive connections themselves form part
of total theory ... But in that case there is nothing for the inferential links between sentences to
consist in. They cannot be replaced by super-inferential links compelling us, if we accept certain
logical principles, to accept also the consequences under those principles of other sentences we
accept for any such super-logical laws could in turn be formulated and considered as sentences no
more immune to revision than any other.'

Dummett's claim, in other words, is that no sense can be made of empirical
pressure forcing re-adjustment in the truth-values of the sentences a
speaker holds true in his language-cum-theory if inferential relations
between those sentences are themselves revisable. To attempt to meet this
problem by replacing those inferential rules by super-inferential rules
brings us no closer to understanding the content of these new rules since
these too are subject to Duhem's Thesis according to Quine. We are thus
precluded from understanding the inferential relations between sentences in

\(^{14}\) loc. cit.

\(^{15}\) 'Frege's Distinction between Sense and Reference' p. 136 in TOE pp. 116-144.

\(^{16}\) 'The Justification of Deduction' p. 309 , loc. cit.
a given language on Quine's theory and thereby prevented from theorising about meaning at all.

I think Dummett's argument is a very powerful one to bring against the global application of Duhem's Thesis to a language. Still, I am not convinced that Quine has no answer to it. It is true that Duhem's Thesis cannot make clear the contents of the logical constants and thus of the contents of theoretical sentences (since these consist in their inferential relations with other sentences). But, I take it that Quine would contend that it is not meant to - this is the task of a naturalised epistemology instead.

Quine undertakes this task in 'The Nature of Natural Knowledge' and more fully in The Roots of Reference. In 'The Nature of Natural Knowledge', for instance, Quine tells 'in outline and crude conjecture' a psychologistic story about: 'how one might start at the observational edge of language and work one's way into the discursive interior where scientific theory can begin to be expressed.'

Thus the child might learn conjunction by noticing that the adult assents to 'p and q' in just those circumstances in which he assents to 'p' and also to 'q'. The route to learning the use of other truth-functions is a little more involved and the route to mastery of the apparatus of objective reference and of variables still more so.

But will this psychologistic story really suffice as an answer to Dummett? Dummett was after a specification of the contents of inference rules, rules which are meant to tell us how we ought to reason; there is nothing in the above account which explains why Modus Ponens is a good inference rule or why affirming the consequent is a bad way to reason. The naturalised epistemologist simply tells us how we all end up more or less converging in

17 'The Nature of Natural Knowledge' p. 77.
18 Compare the account of the acquisition of the relative clause at pp. 76-77.
our inferential habits. Dummett would complain (and this is one of his principal objections to holism) that this psychologistic account fails to show how or why the inferential rules we use are justified.

Quine's probable reply to this criticism would be that the only measure of rationality we possess is that provided by science. We can justify our scientific practices generally and our inferential practices in particular only by using the techniques and doctrines of science itself for it is its successful practice alone that justifies science. Scientific theories can be expressed in classical first order logic\(^{19}\), an idiom which is both simple and which makes ontological commitments manifest; our classical inference rules are therefore justified as part of a revisable but generally successful theory of the world.

I cannot hope to resolve this dispute here\(^{20}\), but we should note in passing that whilst psychologistic theories by themselves invariably look implausible, Quine's advocacy of them is well motivated by his desire to uncover the empirical foundation of our language/theory\(^{21}\).

Quine's empiricist philosophy of language and epistemology are fashioned around the concept of empirical content - the meanings of all sentences depend on their conditioning to observation sentences which are directly keyed to stimulatory patterns, and it is the task of the empirical psychologist to discern these pathways of conditioning from sensory content to assent.

\(^{19}\) 'all traits of reality worthy of the name can be set down in an idiom of this austere form if in any idiom.' Word and Object. p.228.

\(^{20}\) I return to the question of the justification of our inferential practices at 54.6.

\(^{21}\) 'the paths of language learning, which lead from observation sentences to theoretical sentences, are the only connection there is between observation and theory.' in 'The Nature of Natural Knowledge' p.79
In 'A Coherence Theory of Truth and Knowledge'\textsuperscript{22}, Davidson explicitly disavows Quine's 'unassailable' tenets of empiricism - the relation between sensations and beliefs cannot be one of justification, he claims, but is rather a causal one; all that can justify a belief is a further belief; meaning and knowledge are causally dependent on sensory stimuli. Duhem's Thesis is unavailable to Davidson insofar as it depends on the notion of a confrontation between our beliefs and experience. Davidson finds the idea of such a confrontation absurd since: 'we cannot get outside our skins to find out what's causing the internal happenings of which we're aware.'\textsuperscript{23}

We should look elsewhere than Duhem's Thesis for the source of Davidson's holism. Dummett cites the following claim of Davidson's as a 'characteristic expression' of holism at TOE p.134:

'To give up the analytic-synthetic distinction as basic to the understanding of language is to give up the idea that we can clearly distinguish between theory and language. Meaning, as we might loosely use the word, is contaminated by theory, by what is held to be true.'\textsuperscript{24}

These words could just as well have been uttered by Quine, and Dummett goes on to discuss this passage 'in terms of Quine's ideas'\textsuperscript{25}. Quine and Davidson both agree that language and theory (or meaning and belief) are inextricable but for rather different reasons.

For Quine, the inextricability thesis amounts to this: because the empirical evidence only supports sets of sentences without supporting any one individual sentence and because science is simply a refinement of our ordinary talk, the empirical content of both ordinary language and science is inextricably confounded. The radical translator looking for the stimulus conditions that uniquely prompt assent to a given queried sentence will be

\textsuperscript{23} 'A Coherence Theory of Truth and Knowledge' loc. cit. p.312.
\textsuperscript{24} 'On the very idea of a conceptual scheme' loc. cit. p.187.
\textsuperscript{25} loc cit.
thwarted overall. Quine's Inexricability thesis thus depends on two notions Davidson rejects - empirical content (at least on Quine's construal of that notion) and Duhem's Thesis.

Inexricability has a far more straightforward sense for Davidson - whether a native holds a sentence true will depend jointly on his propositional attitudes and on what the sentence means: so one cannot without circularity use the subject's meanings as evidence for his attitudes or conversely use his attitudes as evidence for his meanings.\(^{26}\)

Dummett distinguishes this holism 'in respect of evidence' from a 'holistic view of language'\(^{27}\) - the latter relates to how a speaker's implicit knowledge of a theory of meaning for his language determines the use he makes of its expressions. Dummett attempts to articulate what he takes to be Davidson's form of holism in the following way:

(i) We start with the class of sentences the native holds true, call it \(T\).

(ii) We define a preferred assignment as that which maximises the number of sentences in \(T\) which are actually true.

(iii) We then simultaneously assign references to singular terms and extensions to predicates such that the referent of \(t\) is that object assigned to \(t\) under the preferred assignment and similarly for predicates.

What makes this theory holistic is that it is the same piece of knowledge that is required to grasp the senses of all names and predicates - viz. the make-up of \(T\). Dummett claims that it is beyond human capacities to simultaneously determine the references of all singular terms and predicates in this manner. For it is no help to be told that the referent of \(t\) is that individual of whom the majority of predicates derived from sentences

\(^{26}\) Although it has a straightforward sense, Davidson uses (IBMT), as we saw in S1.4, to draw substantive metaphysical conclusions about the nature of the mind.

\(^{27}\) 'What is a Theory of Meaning?' p.127.
containing it are true if one does not already understand what it is for any one such predicate to be true of any individual. Thus on this version of holism also a systematic theory of meaning is impossible.

Now I agree with Dummett that such a theory lacks any plausibility for just the reasons he adduces, but I think that this could not be what Davidson's holism consists in. Davidson has never held the stochastic version of the Principle of Charity that Dummett saddles him with here. As we saw in §1.4, the spirit of that principle is to optimise truth, not statistically maximise it.

What are the sources of Davidson's holism? Firstly, the methodology for constructing and testing a truth theory itself is holistic - the interpreter cannot do other than pair sentences that the native holds true with sentences he himself holds true under like circumstances and it is only at the end of this process when the theory has shown that it can pass the relevant formal and empirical constraints that the interpreter can have any confidence that the metalinguistic sentences used on the RHS of the T-sentences translate or give the truth-conditions of the object language sentences mentioned on the LHS. The constraints are constraints on the theory as a whole.

There is a further crucial source of holism for Davidson - meaning is itself treated holistically. Dummett is naturally aware of this further element (a 'holistic view of language') and is I think right to see this as surviving a separation of meaning from belief ('holism in respect of evidence'), but I do

28 loc.cit. p.133.

29 It seems very doubtful that a theory in which 'Maximise the number of held true sentences' is an inviolable canon could do the most fundamental thing an adequate theory of interpretation must do - namely, rationalise the actions of the native. A 'preferred assignment' might attribute bizarre reasons for actions to the native consistently with that maxim just because of that, this could not be what Davidson's holism consists in.
not believe he correctly perceives its nature. In order to give content to it he proposes the stochastic version of the Principle of Charity which misconstrues its import.

Davidson is chary of talking of meanings at all and hopes to achieve a means of interpreting the speech of a native by articulating just the truth-conditions of his sentences together with their logical form in a truth theory for his whole language. What makes this theory holistic is the thesis that the inferential connections between sentences can only be justified by a systematic theory of truth for the language, showing that they lead from true sentences to true sentences in every case. Ascriptions of logical form can only be justified in the context of a truth theory for the entire language, the resultant logical relations between sentences affording the only genuine criterion of when a language commits us to the existence of entities.

Now it is true that on this holistic theory, the meaning of a statement: 'simply consists in the place which it occupies in the complicated network which constitutes the totality of our linguistic practices.'

But without some explanation of the crucial connection between the enterprises of articulating logical form and of discerning ontological commitments which the construction of a truth theory for a language forges, this metaphor lacks any clear motivation.

Moreover, it is essential to realise that for Davidson, sentences still have objective truth-conditions of their own, unlike Quine’s picture in which most sentences lack individual verification conditions and truth is merely a disquotational property. Indeed, were it not for the fact that sentences are lawfully connected with states of affairs in the world, interpretation could not even begin.

30 ‘The Philosophical Basis of Intuitionistic Logic’, loc. cit. p.218
Because he rejects the notion that we understand a sentence by testing its truth either singly or in concert with other sentences (as in Duhem's Thesis) against reality, experience, sensory stimuli or any uninterpreted empirical content, Davidson eschews 'confrontationist' models of our grasp of a statement's truth-conditions. Rather, we understand the truth-conditions of a given sentence when we discern the logical form of that sentence and thereby appreciate the inferential relations between that sentence and others of the language in a truth theory for the language as a whole.

Davidson subscribes to three holistic theses in effect -
(1) Sentences derive their contents from their place in a truth-theoretic inferential network;
(2) Beliefs similarly derive whatever content they have from their inferential and evidential relations with other beliefs in a system of beliefs;
(3) Our ascriptions of beliefs and meanings to speakers are inextricably interlocked.

Dummett's central objections to holism are these:
(1) Holism precludes the possibility of a systematic theory of meaning since:
   (a) The inferential relations between sentences cannot be made out on such a theory.31
   (b) The appearance of structure is a sham in a holistic theory as there is no distinctive behavioural manifestation of the grasp of any propositions of the truth theory.32
   (c) A meaning theory based on a holistic view can give no content to the notion of a speaker's being mistaken about the meaning of an expression.33

(2) Holism cannot account for the usefulness of deductive reasoning, nor can it even consider how such reasoning might be justified.34

(3) Holism cannot account for the progressive acquisition of a language.35

31 Frege pp. 596-597.
32 'What is a Theory of Meaning?' p. 116, p. 121; TOE p. 136.
33 'What is a Theory of Meaning?' p. 116-119.
34 TOE pp. 303ff; 'What is a Theory of Meaning (II)?:' p. 105.
35 Frege p. 599, p. 622; TOE p. 177; 'What is a Theory of Meaning?' p. 137, others.
Now I have already queried Dummett’s claim that Quine cannot answer 1(a), and one of the apparent strengths of Davidson’s position is that it precisely does explicate the inferential relations between sentences of a language by means of a truth theory for that language.

1(b) alleges that the appearance of structure in a holistic theory of meaning is a fraud. For Dummett a theory of meaning is a theoretical representation of the practical ability to use the language in question, this practical ability being cognitively based, consisting in implicit knowledge of the propositions of the theory of reference. The theory of sense has the crucial responsibility of pairing practical abilities with known propositions of the theory of reference.

A ‘Full-Blooded’ theory of meaning includes a theory of sense which discharges its function properly by pairing known propositions of the theory of reference with practical abilities which manifest a speaker’s understanding of the propositions in question. But a ‘Modest’ theory of meaning abrogates its responsibility by just insisting on a theory of reference without a theory of sense. The appearance of structure in a ‘modest’ theory of meaning is therefore a sham since the structure corresponds to no cognitive ability on the part of those speakers whose competence it is supposedly modelling.36

Dummett thinks the vacuity of modest theories of meaning is no more in evidence than in their recursive clauses for the logical constants. Such theories do not explain the senses of the logical constants because they use those very constants in the metalinguistic stipulations of their meanings in the truth theory for the object language. The logic discerned in the object

---

36 Dummett initially charged that Davidson’s holistic theory was ineradicably modest (‘What is a Theory of Meaning?’ p.102) but later retracted this upon reflection (p.128), offering the stochastic version of the Principle of Charity as Davidson’s way of avoiding the charge of modesty.
language will be just that which the metalanguage reads into it; but no
justification is given for the logic of the metalanguage.37

Dummett argues at 'What is a Theory of Meaning? (II)' p.108 that:
'A trivial axiom for any expression, whether logical constant or any other kind, does not, in itself,
display in what an understanding of the expression consists, but throws the whole task of
explaining this upon the theory of sense, which specifies what is to be taken as constituting a grasp
of the proposition expressed by that axiom.'

Dummett's thought is that just because it eschews a theory of (Dummettian)
sense, and only provides 'trivial' axioms for L's expressions, a modest theory
of meaning is debarred from giving any account of what an understanding of
these expressions consists in.

It seems to me that both 1(a) and 1(b) present serious challenges to the
Realist proponent of a holistic theory of meaning. So too does 1(c), the charge
that a holistic theory can give no content to the notion of a speaker's being
mistaken about the meaning of an expression. A holist might be tempted to
regard this with equanimity as a natural consequence of the inextricability
of belief and meaning (IBMT), but I think it is a highly undesirable
consequence.38

When we settle for Davidson's 'bootstraps' methodology of attributing
coarse-grained attitudes to the native which we seek to continuously refine
in the light of further evidence, but take a more Realistic attitude to the
mental life of the native than Davidson does, allowing that, under certain
unusual conditions, the native could be massively in error in his general

37 There is no compunction to read the constants as having classical meanings because precisely
the same recursive clauses would feature in a metalanguage which employed a non-classical logic
- one could thereby endow the object language operators with non-classical meanings by adopting
'trivial' disquotational axioms (i.e. ones of the form "snow is white" is true iff snow is white) in
such a metalanguage ('What is a Theory of Meaning? (II)' p.107).
38 If a speaker's use of an expression is intentional at all, it must be possible for the speaker to
fail in his intention to use the expression correctly; a theory of meaning should, accordingly, give
some account of how such errors are possible.
beliefs about the world, we have, I contend, made room on a holistic theory of meaning for the possibility of error and ignorance on the part of those whose speech we seek to interpret.\textsuperscript{39}

Dummett's second main criticism of holism - that it cannot account for the usefulness of or provide a justification for, deduction - is examined obliquely in §4.6, the conclusions reached there being, so far as I can see, acceptable to a holist.

As to Dummett's third criticism, that holism cannot account for a child's piecemeal acquisition of language since it cannot make sense of the child's knowing part of a language, I wonder whether Dummett's description of the child's language-learning situation is not tendentious from the holist's point of view. Presumably, the holist will see the child as at any stage of the language-learning process possessing a complete, albeit primitive, language which may be a proper sub-part of 'our' language, but will most probably only approximate to such. Language-learning involves the (surprisingly swift) acquisition and abandoning of a series of progressively less primitive languages, each stage growing out of its predecessor by an evolutionary process involving revisions, 'paradigm changes' and so on with respect to that predecessor. We might choose to describe this process as the child's knowing at any one time only part of English, say, but this is misleading.

\textsuperscript{39}Dummett argues: 'if a member of the linguistic community holds a divergent theory of truth for the language, he will tend to diverge more in his judgements than most speakers do from the majority. But, since no finite set of such divergences will, in itself, reveal his reliance on a non-standard theory of truth, it is hard to see how either he or the other speakers or we as observers could ever detect this, or how, once discovered, it could be corrected.' (What is a Theory of Meaning? p. 117) If Davidson just acquiesces in this result, it becomes very doubtful how he can avoid Dummett's charge at loc. cit. p. 118 that his theory of meaning is of necessity solipsistic. Davidson seems to have drawn this conclusion (or something very like it) in 'A Nice Derangement of Epitaphs', in Grandy, R and Warner, R (eds.) Philosophical Grounds of Rationality: Intentions, Categories, Ends Clarendon Press, Oxford, 1986, pp. 157-174, but it seems to me that in that article Davidson has misconstrued the significance contextual factors play in interpretation.
inasmuch as it suggests that what is really going on is that the child possesses, at any stage, only a partial language. At some stage he will have mastered English, to the extent that we are willing to attribute such mastery to any person, 'without it ever being accurate to say he has mastered part of it first.'

Davidson ‘Replies to Essays X-XII’ p. 252 in Hintikka, M and Yermazen, B (eds.), loc. cit. I should quote the surrounding context of this remark. Davidson is considering Patrick Suppes’ Dummett-styled objection that a child could never learn his native tongue at all if he (Davidson) is correct about the inextricability of belief and meaning. Thus he says: ‘I do not really think I must produce “the details of an actual theory of language acquisition and cognitive development” in order to answer this question (i.e. how language learning is possible in the light of (IBMT)) for there is no reason a child cannot slowly master a complex system without it ever being accurate to say he has mastered part of it first. The trouble with this answer is that it leaves us not knowing how to describe the early stages, just as the general thesis that animals don’t have beliefs leaves us without our usual way of explaining and describing their behaviour.” I agree with Suppes and Dummett against Davidson that (IBMT) construed in Davidson’s way as a substantive metaphysical thesis about the nature of belief and meaning does make it extremely hard to see how anyone, let alone a child, could come to grasp a language. But I disagree with Dummett that the solution to this is to abandon holism about meaning. I think the solution is to retain the aforesaid holism but balance it with a more realistic appraisal of the mental so that (IBMT) simply records the fact that the mental is itself holistic. Moreover, I think that Davidson concedes too much to the opponent of holism by claiming that semantic holism ‘leaves us not knowing how to describe the early stages’ – we describe them as in the text above.
§3.1 TRUTH, ASSERTION AND ASSERTIBILITY

§3.1.1 Dummett on Assertion

Dummett’s project is to exhibit truth as a construct out of the primitive notion of correct assertibility.

At Frege p.3 Dummett claims that it ‘belongs to the essence of language’ that:
‘assertions are understood as governed by the convention that the speaker is aiming at uttering those whose truth condition is fulfilled.’

Yet if the concept of truth derives from the primitive notion of an assertion’s being correct (or incorrect), the claim above appears to be a truistic one since the convention is then just to utter those statements it would be correct to assert. For this reason, it seems very unlikely that Dummett intends to characterise assertoric force by means of a convention to utter true sentences. So what does Dummett take assertoric force to be?

Dummett considers and dismisses the general possibility that illocutionary forces are individuated by the intentions of speakers: intention cannot bear the weight of the distinctions, since the relevant intentions can only arise against the background of a general convention endowing certain utterances with a certain significance.1 So instead of starting with utterances as an undifferentiated class and attempting to specify which constitute assertions, which questions etc, by reference to the intentions of the speaker:
‘the correct approach is to consider utterances as conventionally demarcated into types, by means of the form of linguistic expressions employed, and enquire into the conventions governing the use of the various types of utterance.’2

1 Frege p.300.
2 loc. cit. p.302.
Convention is thus invoked to characterise the illocutionary force of the linguistic act of assertion. Each distinguishable type of illocutionary act is governed by a distinctive sort of convention - by those conventions, we can correlate illocutionary acts with distinctive purposes on the part of the speaker, pairing assertions with intentions to utter truths, commands with intentions to get hearers to realise truths, questions with intentions to find out about truths and so on; the intentions in question being those that the speaker is conventionally taken to have via his issuing utterances with the appropriate force.

As to what signals that an utterance has one rather than another type of force, convention is again asked to provide the explanation - Dummett thinks it 'roughly correct' that the indicative mood carries assertoric force and similarly for the relation between other moods and forces.3

Dummett does not think it arbitrary that the person who asserts that p is thereby taken to have the intention to utter a truth. In fact he clearly thinks there is a conceptual connection between the assertion and the relevant intention. For he dismisses as 'spurious' the possibility that speakers could have adopted the convention to utter falsehoods in issuing statements with assertoric force, acknowledging Wittgenstein as making a 'cognate point' at Tractatus 4.062.4

Wittgenstein's argument there was that if a speaker uttered 'It is raining' and meant by it that it was not raining, then since truth depends upon meaning we should say that what he said was true not false if in fact it were not raining -

3 loc. cit.p.303.
4 loc. cit.p.318.
If by "p" we mean "¬p" and things are as we mean them, then "p" is on the new interpretation true and not false.

So it must be strictly incorrect on Dummett's view to say that the intention to utter a truth is conventionally related to an assertion, given that an alternative regularity to utter falsehoods (which seems just as useful) is 'spurious'.

Dummett cannot be wanting to characterise the convention governing assertion primarily in terms of the speaker's being imputed a certain intention or belief. Now very crudely there are two possible ways in which assertoric force could be characterised:

(1) By means of the beliefs or intentions on the part of the speaker.

(2) By means of the commitments that a speaker makes in asserting a sentence.

In accord with (2), there is at least one well-defined commitment which the assertor makes to his audience - namely to withdraw his assertion that p if he cannot, when challenged, supply adequate grounds for believing p.

Jim Mackenzie has presented a model of the rules of a rational dialogue between two participants which could be used to characterise assertoric force quite precisely. The model has the following structure:

---

(I) 

| Statements | \( p \) | \( \neg p \) |
| Questions | \( ?p \) | Is it the case that \( p \)? |
| Withdrawals | \( \neg p \) | I'm not sure that \( p \) |
| Challenges | \( \neg p \) | Why is it to be supposed that \( p \)? |
| Resolution demands (sets) | \( \neg p \) | Because \( p \) |

(II) Commitment Rules

CR\(_I\) Initial commitments are null.

CR\(_Q\), CR\(_Q\) Questions and Resolution demands do not affect commitments.

CR\(_S\) After a statement \( p \), \( p \) is added to the commitments of both speakers.

CR\(_W\) After a withdrawal \( \neg p \), \( p \) is removed from the speaker's commitment; the hearer's commitment is unaffected.

CR\(_Y\) After a challenge \( \neg p \), \( p \) is removed from the speaker's commitment. \( \neg p \)

is added to the speaker's commitment, and \( p \) is added to the hearer's commitment.

CR\(_G\) After a ground \( \neg p \) in reply to the challenge \( \neg p \), \( p \) and \( \neg p \) then \( q \) are added to both commitments.

(III) Rules of Dialogue

RlogAx: Nobody may deny, challenge, or withdraw an inference.

RQuest: After \( B \) asks \( ?p \), \( A \) must reply with \( L \) where \( L \) answers \( ?p \).

RChall: After \( B \) challenges \( \neg p \), \( A \) must reply with either:

(i) \( \neg p \)

or (ii) \( r \rightarrow p \), where \( B \) is committed to all of \( r \) and \( p \) is a logical consequence of \( r \).

or (iii) \( \neg r \) where \( B \)'s commitment contains both \( q \) and if \( q \) then \( p \).

RGround: \( A \) may say \( \neg p \) only if the preceding event was \( B \)'s challenge \( \neg q \).

RRes Set: \( A \) may demand \( r \rightarrow p \) only if \( B \) is committed to all of \( r \) and \( p \) is inconsistent with a statement \( q \) in \( B \)'s commitment.

RSetRes: After \( B \) demands \( r \rightarrow p \), \( A \) must reply \( \neg p \) where \( p \in \Gamma \).

RResCon: \( A \) may demand \( r \rightarrow p \) only if \( B \) is committed to all of \( p \), together with \( p \)'s being a logical consequence of \( r \) and the preceding event was \( B \) saying either \( \neg p \) or \( \neg p \).

RConRes: After \( B \) demands \( r \rightarrow p \), \( A \) must reply either \( p \) or \( \neg q \) where \( q \in \Gamma \).

Now I have omitted from the model the all-important accepted inferential rules. But my purpose was simply to give some indication as to how the
notion of assertoric force could be defined in terms of speaker-commitments incurred in the course of a dialogue between two speakers.

The characterisation of assertoric force offered by (Z) will be attractive to anyone hostile to psychologism. Dummett shares Frege's distrust of psychological incursions in the theory of meaning:

"The procedure we have agreed on, however, is to give an account of these exterior linguistic acts directly, by describing the conventions governing the use of the various forms of expression: the alternative procedure, of describing first the nature of the various interior acts, and then explaining the various types of utterance as being conventionally used to express the occurrence of these mental acts, we have rejected as leading to the confusion of psychology with logic." 6

On Dummett's view, this latter confusion leads to the mistaken picture of assertion as the manifestation of a mental act. 7 After adverting to Russell's attempts to distinguish a logical from a psychological sense of 'assertion' and to the early Wittgenstein's counter that 'assertion is merely psychological',

**Dummett remarks:**

"This supposed 'psychological' kind of assertion which appears in Russell and Wittgenstein is a phantasm produced by the mistake of interpreting assertion as the manifestation of an internal mental attitude adopted toward the proposition; in fact, there is nothing any more psychological about assertion than about the sense expressed by a sentence." 8

The natural response to this is that if assertion is not a psychological act, it is hard to see what else it could possibly be. But Dummett's attack is directed against psychologism - which is the view that an inner process of assent to a proposition must precede the public act of assertion which merely exteriorises it; a view dependent on the model of language as a code for our pre-existent and perfectly determinate thoughts.

I agree that the psychologistic model above is of dubious coherence, but I do not think that we need accept it in order to make out the connection

---

6 ibid p.313.
7 ibid p.314.
8 ibid p.312, italics mine.
between assertion and belief.\textsuperscript{9} Dummett thinks that we have only two choices - to describe the conventions governing assertion directly or to accept psychologism. Having rejected psychologism, we must explain what it is to judge a thought to be true as 'the interiorisation of the act of assertion'\textsuperscript{10}.

'We have opposed throughout the view of assertion as the expression of an interior act of judgement; judgement, rather, is the interiorisation of the external act of assertion. The reason for viewing the two this way round is that a conventional act can be described, without circularity, as the expression of a mental state or act only if there exist non-conventional ways of expressing it.'\textsuperscript{11}

---

\textsuperscript{9} Thus we can agree that there need have been no \textit{event} which was S's judging that \(p\) temporally distinct from the event of S's asserting that \(p\), though S's judgement could have preceded his giving expression to it through his assertion. So Dummett is right to oppose the psychologistic model on which assertion is to be understood as an encoding of an already determinate and temporally prior judgement, which judgement is in all cases to be construed as the cause of the agent's assertion at any given time. But it does not follow that there are no mental causes of an agent's assertion or that judgement has to be viewed behaviouristically as the 'interiorisation' of assertion (saying in one's heart?). According to Davidson's position on intentional action adumbrated in S1.1.2, assertion is the expression or communication of a judgement, which judgement may, but need not, be a causal factor in the agent's production of the assertion. What will always be the main causal determinant in acts of asserting that \(p\) will be the agent's beliefs (or judgements) from which, in conjunction with all his other concurrent attitudes and evidence, he reasons to \(p\), which belief (or judgement) he gives expression to in asserting that \(p\).

\textsuperscript{10} loc. cit. p.362.

\textsuperscript{11} op. cit.
It is opaque to me how one could 'interiorise' the act of assertion. But be that as it may, Dummett presents us with a dichotomy between conventions and intentions without himself providing any analysis of what a convention might be. Yet there is at least one highly plausible analysis of the notion of convention which rejects such a dichotomy - namely David Lewis's theory of convention. On Lewis's theory, a behavioural regularity $R$ arises as a solution to a co-ordination problem and persists as a convention amongst a populace $P$ when the preference of members of $P$ is to perform $R$, conditional on their expectation that all or most other members of $P$ will perform $R$. The convention will only persist in $P$ if members of $P$ are mutually aware of each others intentions, Lewis argues.

Lewis's analysis would seem unavailable to Dummett. For as we have seen, on Dummett's theory an intention to utter true sentences cannot itself be used to demarcate assertoric force; this has to be done by 'describing the conventions directly'. But if Lewis is right, as I believe him to be, we will only be able to explain why a convention to utter true sentences should persist in a populace at all by appeal to the intentions and beliefs of speakers - although it might be preferable to characterise assertoric force by means of the commitments a speaker incurs in the course of dialogue, the explanation of what those commitments consist in will have to advert to the fact that one who utters $p$ assertorically seeks thereby to convey his judgement that $p$ to his audience.

---

§3.1.2 Truth and Assertibility

The Realist believes that there is a simple relation between truth and assertibility - a sentence is assertible just in case there are good grounds for believing it to be true, where the relevant notion of truth is classical truth. For the Anti-Realist, the relation is less straightforward - truth, if it is a coherent concept at all, can only be some construct from the more primitive notion of correct assertibility.

Dummett argues that assertion is an act which can only be evaluated as correct or as incorrect. There is no third possibility - that is, no assertion can be neither correct nor incorrect (the principle he calls Tertium Non Datur (TND)). Now whatever notions of truth and falsity we might think we can give substance to, they must have their origins in our primitive and foundational practice of evaluating assertions as correct and incorrect, on pain of their failing to connect with that practice at all. In so evaluating assertions, we learn to distinguish the grounds a speaker has for making an assertion from the objectively agreed truth-value of what he states. Thus we arrive at the concept of truth as correspondence with Reality by learning to make and answer criticisms to do with the grounds a speaker has for his assertions from criticisms to do with the truth value of the content of that assertion. Hence we could identify $s_k$'s being true with its being objectively correct.

Thus far there is nothing in our practice of issuing and criticising assertions to justify the notion of truth as an epistemically unconstrained correspondence between our assertions and Reality, one that may obtain independently of whether we can recognise it or not. So to the extent that we think of the sense of a sentence as the content of its assertion, Dummett
argues, there can be no justification for allowing our declarative sentences to take any value other than truth (objective correctness) and falsity (objective incorrectness).

Yet in order to give a systematic account of the use of more complex sentences, we might well be forced to introduce more than one truth-value or possibly even truth-value gaps. But then we would be obliged to show how these assignments mesh with the primitive evaluations of assertions as correct and incorrect. This would involve allowing that there are different ways in which an assertion can be correct or incorrect.

Dummett gives several examples of how this might work: for instance if we wished to preserve the semantic principle that a sentence is false iff its negation is true, we might have to allow that an atomic sentence, \( s_k \), containing a vacuous singular term took a truth-value other than truth or falsity, call it \( U \) ('undecided'). We could then give a truth-functional account of the '¬' operator. But the mere knowledge that \( s_k \) took the value 'true' in A-type circumstances, the value 'false' in D-type circumstances, and the value 'undecided' in N-type circumstances, would not suffice to impart a knowledge of the sense of \( s_k \). In order to understand the content of an assertion of \( s_k \), we would have to at least know:

'which states of affairs are understood to be ruled out by making such an assertion; if none of those states of affairs obtains, then the assertion is correct.' \(^{13}\)

\(^{13}\) loc. cit. p.347
If we do this for the case of a sentence containing a vacuous singular term, we will see that the speaker who asserts 'St Thomas More's eldest son became a protestant' must be ruling out the possibility that More had no sons and would retract it if he discovered that he in fact did; if he had meant to allow for this possibility, he would have said 'If St Thomas More had any sons, the eldest became a protestant.' So, we must regard the value $U$ as registering another way for the assertion of an atomic sentence $s_k$ to be incorrect. Adherence to the Negation Principle (NP) that $s_k$ is false just when $\neg s_k$ is true gives us a justification for calling atomic sentences with vacuous names 'undecided' or 'neither true nor false' - namely, that we can then construe other sentences as their negations. Dummett concludes that: 

'... the linguistic act of assertion makes ... no intrinsic provision for the introduction of a gap between two types of consequence which the making of an assertion might be supposed to have. ... the rationale ... for the introduction of a gap ... always relates, not to the use of the sentence by itself to make an assertion, but to its use as a constituent of complex sentences, in particular, of sentences formed by the application to it of a negation operator.'

14 loc. cit. p.345.
15 loc. cit. p.347.
16 loc. cit.
Thus, Dummett believes that nothing in the practice of assertion justifies attributing to speakers a classical notion of truth - that none of the evidence so far adduced forces us to depart from the equation of truth with correct assertibility. We may have to distinguish different ways in which an assertion might be true (correct) or false (incorrect) in order to give a satisfactory description of our uses of sentential operators but this just forces us to operate with a more sophisticated notion of correctness. Likewise, the distinction between the grounds a speaker has for his assertions and the objectively agreed truth-value of what he says, while it might provide some impetus for thinking of truth as a correspondence between a statement and Reality, does not justify the notion of that correspondence being epistemically unconstrained - of its being possible that $s_k$ might be true even though no-one could in principle determine its truth. Dummett thinks the Realist notion of truth arises as a consequence of the semantic behaviour of sentences as antecedents of indicative conditionals and of tenses and moods within the scope of sentential operators. Thus he writes:

"The notion of truth, as opposed to justifiability, that may be induced by the use of a sentence as the antecedent of a conditional, or as modified by a tense-operator, remains one for which the law of bivalence strictly holds; these considerations merely allow us to distinguish between having a good reason for saying something and being right in saying it. In the absence of such a distinction, however, we become uncertain how to apply the notion of truth at all: it is precisely because indicative conditionals of natural language can neither be negated nor appear as antecedents of more complex conditionals that philosophers have floundered over the application to them of the notions of truth and falsity. It is, in fact, from the distinction between truth and justification that the realistic conception takes its rise. One reason why such a conception appears so plausible is that the notion of truth is born in the first place from the necessity to distinguish between it and the epistemic notion of justifiability: and this necessity is in turn imposed by the requirements for understanding certain kinds of compound sentence."

How precisely do the constructions mentioned above give rise to the classical conception of truth? Dummett writes:

17 loc. cit.p.451, italics mine.
"If future-tense sentences could not come within the scope of sentential operators, there would be no place for such a distinction between justification and truth. We should, for example, have no basis for distinguishing between an expression of intention and a statement of intention, that is, between the forms ‘I am going to marry Jane’ and ‘I intend to marry Jane’, which differ, not in respect of the circumstances in which their utterance is justified, but solely in their truth-conditions. This distinction has to do solely with the different behaviour of the two forms as constituents of more complex sentences, and, particularly, as antecedents of conditionals."

Following Robert Brandom, let us call the compounding devices that Dummett claims induce a separation between assertibility and truth, Truth-Inducing-Sentential-Contexts (TISC’s). Our question is whether the classical conception of truth is forced upon us by acknowledging the existence of TISCs. Dummett in Frege seems to think it is, for he wrote, as we saw that:

‘The notion of truth, as opposed to justifiability, that may be induced by the use of a sentence as the antecedent of a conditional, or as modified by a tense-operator, remains one for which the law of bivalence strictly holds.’

Certainly, if truth must be opposed to justifiability in such TISC’s, a classical conception alone seems able to fulfil the role of the auxiliary semantic concept required to explicate the assertibility conditions of the resultant compound sentences.

But it is by no means obvious that this is so.

Let us examine then whether tense-operators induce any distinction between assertibility and truth. Dummett takes this question up at Frege

---

18 loc. cit. p. 450.

19 ‘Truth and Assertibility’ The Journal of Philosophy, LXIII, (1976), pp. 137–149. Examples of such TISCs are constructions involving ‘believes that’, ‘it is possible that’, ‘it will be the case that’, ‘were it the case that... then...’. If we start with two sentences with the same assertibility conditions p and q, then a TISC will generate two compound sentences that differ in their assertibility conditions.

20 Brandom’s view is more moderate. He points out that for the purposes of the theory, truth is whatever auxiliary concept is necessary to explicate the TISC’s and this clearly underdetermines the actual truth-conditions, hence truth concept, that will be assigned the resultant compound sentence.
There he examines a variety of semantic systems adequate for the formalisation of tensed statements in which the basic semantic notion is not truth at a time \( t \), but truth at \( t \) under a possible total course \( C \) of world history. Each of these semantic systems has a tree-like structure, the branches of the tree representing differing total courses of world history possible at the origin of the universe.

If we call a sentence without temporal operators a **radical**, we stipulate that each radical is determinately true or false at a node, where a node is to be thought of as representing a state of affairs on a particular day, under a total course of world history. Tensed sentences are represented as comprised of a token-reflexive temporal operator and a sentence-radical.

Letting the unit of time be a day, we write 'It was the case \( n \) days ago that ...' as '\( Pn \)'; 'It will be the case in \( n \) days ...' as '\( Fn \)'. A semantics which validates all the laws of classical logic will have such clauses as:

1. **(6)** \( A \) is true (false) under \( C \) iff it is true (false) at every time \( t \) under \( C \).
2. **(7)** \( FnA \) is true (false) at \( t \) under \( C \) iff \( A \) is true (false) at \( t + n \) under \( C \).
3. **(8)** \( PnA \) is true (false) at \( t \) under \( C \) iff \( A \) is true (false) at \( t - n \) under \( C \).

For someone who believes the future is essentially indeterminate, this semantics has two drawbacks, according to Dummett:

'Since it validates all the laws of classical logic, it is impossible to refute the argument for fatalism by rejecting certain of those laws, as applied to statements about the future. ... Secondly, although no one future course of world history is taken to be the actual one, the understanding of the future tense (or of temporal operators relating to times later than the present) depends essentially upon the conception of a possible total future course of history.'

Dummett then suggests an alternative non-classical semantics in which truth is defined in terms of assertibility. He introduces the notion of the coincidence of two total courses of history and uses this to define assertibility at a time \( t \):

---

21 Frege p. 393.
(a) We say that $C$ coincides with $C'$ up to $t^*$ just when for every radical $A$ and every $t^* \lt t$ $A$ is true at $t^*$ under $C$ iff $A$ is true at $t^*$ under $C'$.

(b) $A$ is assertible at $t$ under $C$ iff for every $C'$ which coincides with $C$ up to $t$, $A$ is true at $t$ under $C'$.

If we then identify truth with assertibility thus defined, we obtain a non-classical semantics in which:

(9) $F_n A$ is true at $t$ under $C$ iff, for every $C'$ which coincides with $C$ up to $t$, $A$ is true at $t + n$ under $C'$

(10) $P_n A$ is true at $t$ under $C$ iff, for every $C'$ which coincides with $C$ up to $t$, $A$ is true at $t - n$ under $C'$.

This 'Aristotelian' semantics differs sharply from the classical one in that $A$ may be true at $t$ under $C$ though $F_n A$ is false at $t - n$ under $C$. As a result, although $F_n(P_n A)$ is always equivalent to $A$, $A$ does not imply $P_n(F_n A)$ since $A$ might be true at $t$ under $C$ whilst $P_n(F_n A)$ is false since for some $C'$ coinciding with $C$ only up to $t - n$ but not $t$, $A$ is false at $t$ under $C'$. Dummett comments that this failure of equivalence between $A$ and $P_n(F_n A)$ might be considered 'a weakness of the system'.

He seeks to secure this equivalence in the following way:

Suppose that there is no distinction between the assertibility conditions of future-tensed statements and their truth-conditions - that an assertion of a future-tensed sentence is correct just in case the future event described does in fact take place at the appropriate later time. Then we can construct a non-classical semantics validating the equivalence of $A$ and $P_n(F_n A)$ using the previous non-classical semantics if we but relativize the truth of a tensed sentence not only to its time of utterance but also to its time of assessment. We would then have:

---

22 loc. cit. p.394.
(11) A, uttered at $t$, is true at $t^*$ under $C$ iff for every $C'$ which coincides with $C$ up to $t^*$, $A$ is true at $t$ under $C$.

(12) $F_nA$, uttered at $t$, is true at $t^*$ under $C$ iff, for every $C'$ which coincides with $C$ up to $t^*$, $A$ is true at $t + n$ under $C$.

(13) $P_nA$, uttered at $t$, is true at $t^*$ under $C$ iff, for every $C'$ which coincides with $C$ up to $t^*$, $A$ is true at $t - n$ under $C$.

Dummett notes that the desired equivalence between $A$ and $P_n(F_nA)$ is now restored. The equivalence failed before because it was possible for the courses of world history $C$ and $C'$ to diverge up to the time of assessment. But this possibility is ruled out in this new semantics.

Thus if we were now to evaluate the truth of a Greek soldier's present tensed claim 'A sea battle is being fought off the shores of Pylos now' as uttered in 234 B.C. and compare it with a Greek historian's endorsement of the Delphic Oracle's prediction in 334 B.C. that this very battle would take place one hundred years on, as expressed by the historian's response to the soldier 'It was the case one hundred years ago that a sea battle would take place off the shores of Pylos now', we can use our clauses to show that the soldier's and historian's statements have the same truth-conditions.

For by clauses (12) and (13), 'It was the case one hundred years ago that a sea battle would take place off the shores of Pylos now' is true now in 1986 in the actual course of world history just when for every $C'$ coinciding with the actual course of history up to now, had 'A sea battle is being fought off the shores of Pylos now' been uttered, it would have been true 2220 years ago in $C$. But these are precisely the same truth-conditions for 'It was the case one hundred years ago that a sea battle would take place off the shores of Pylos now' uttered and evaluated in the very same situation as 'A sea battle is being fought off the shores of Pylos now'.

Dummett writes:
we should have no need of the distinction between the genuine future tense, yielding a statement true or false according to what later happens, and the future tense expressing present tendencies, as occurring in, e.g., "The wedding announced between ... and ... will not now take place". The difference between the two uses of the future tense is registered only in compound sentences, such as a conditional whose antecedent is a future-tense sentence, or one involving a compound tense like "was going to ... "

I think that Dummett means to be saying that we grasp the distinction between statements which, though future-tensed, only record present dispositions and statements that really are about the future (such as the Delphic Oracle's prediction above) by being taught to regard certain future-tensed statements uttered in the past as equivalent to certain present-tensed statements (e.g. as in the example above 'It was the case one hundred years ago that a sea battle would take place off the shores of Pylos now' and 'A sea battle is being fought off the shores of Pylos now'). We are taught that this is a mark of the distinction between statements really about the future and future-tensed dispositional statements.

But on an Anti-Realist semantics for tensed statements such as the one with (12) and (13) as its clauses, there will be no distinction between truth and assertibility - a future tensed statement will be both correctly assertible and true just when the future event it describes does occur at the appropriate later date. So we can explain how we could have arrived at a bogus notion of truth transcending all possibility of verification by means of a mistaken construal of our linguistic practices with tensed statements, at the same time correctly construing those features of our practice (such as the semantics of conditionals and tensed statements) which gave rise to this misconception.

If this is Dummett's argument, it is an ingenious one indeed. But even if Dummett and Brandom were right about the origin of the notion of truth in

23 loc. cit. p. 450.
the behaviour of certain sentential operators, TISC's, particularly temporal contexts, the mere project of explicating the assertibility conditions of compound sentences with the aid of truth-conditions ought not worry the Realist.

In order to pose any threat to Realism, the Anti-Realist must show not that the assertibility conditions of complex sentences depend not just upon the assertibility conditions of their sentential constituents, but that the truth-conditions of those compounds do not just depend upon the truth-conditions of their constituent sentences.

For if the truth of a complex sentence does not just depend upon the truth of its atomic sentential parts, the semantic value of a sentence could not be its truth-value and the way a sentence is determined as true could not be constitutive of its meaning, which is precisely the claim of Anti-Realist semantic theorists. This crucial claim still awaits vindication.
§3.2 MODEST AND FULL-BLOODED THEORIES OF MEANING

§3.2.1 Dummett's requirements on an adequate theory of meaning

Dummett contends that an adequate theory of meaning must be 'full-blooded' rather than 'modest'. A modest theory simply states the meanings of the expressions of a speaker's language without providing an account of what that knowledge consists in or how it is manifested in behaviour; a full-blooded theory however does provide such an account.

Dummett introduced the notions of modesty and full-bloodedness at 'What is a Theory of Meaning?' pp 101-2. He writes:

'Hence, if a theory of meaning is a theory of understanding ...(it) must explain what it is to have the concepts expressible by means of that language. ... since these concepts may be grasped by someone who is quite ignorant of that particular language, but knows another language in which they are expressible ... the theory of meaning must ... show or state which concepts are expressed by which words. And an alternative view will be that it is only the latter task which properly belongs to the theory of meaning; that to demand of the theory that it should serve to explain new concepts to someone who does not already have them is to place too heavy a burden upon it, and that all that we can require of such a theory is that it give the interpretation of the language to someone who already has the concepts required.'

Thus, a modest theory gives the interpretation of a language $L$ only to someone who already has the concepts corresponding to $L$'s expressions whilst a full-blooded theory can be used to explain the requisite concepts to someone who does not already have them.

---

1 loc cit; italics mine. McDowell notes the slide from explaining concepts to a novice to explaining what it is to have those concepts at 36n p. 165 of 'On the Sense and Reference of a Proper Name' loc. cit. in Platts, M (ed) Reference, Truth and Reality, op. cit.
I find these distinctions unclear and their point hard to discern. Any adequate meaning theory must state the meanings of the expressions of the object language and in so doing will associate concepts with words. If it is to be used to communicate these meanings to an agent, it must be formulated in a language that the agent understands, so to that extent the communicator must presuppose that the agent possesses certain concepts — just those concepts (or a significant subset thereof) that the words of the language in which the theory is formulated are associated with. Insofar as it does this, then, any theory of meaning used to communicate the meanings of the expressions of an object language L will be constrained into modesty. An unconstrained full-blooded theory of meaning would then seem to be incommunicable.

Now as McDowell observes, Dummett does come very close to demanding the impossible — that a meaning theory be stated in no language at all at all.

"What is a Theory of Meaning?" pp.103-104, for he writes:

"Thus a translation manual presupposes a mastery of some one other language — that into which the translation is made — if we are to derive from it an understanding of the translated language; but a modest theory of meaning presupposes a mastery of some, though unspecified, language, if we are to derive from it an understanding of the object-language. The significant contrast would, however, appear to be not between a theory which (like a translation manual) makes a specific presupposition and one which (like a modest theory of meaning) makes as heavy a presupposition, though less specific; but between theories which (like both of these) rely on extraneous presuppositions and those which (like full-blooded theories of meaning) involve no such presupposition at all."

2 For example, it is puzzling that Dummett should require that the relevant contrast be between a theory that is only intelligible to someone who possesses certain concepts and one intelligible to someone who does not possess those concepts, since there is no obvious connection between that requirement and the earlier contrast between theories that explain what possession of the concepts expressible by a language consists in and theories that do not.


4 Of course it would be utterly perverse to attempt to teach a language to someone by providing him with a theory of meaning for it — as perverse as attempting to teach someone to ride a bike by getting him to solve the set of differential equations corresponding to the requisite bodily movements.

5 'On the Sense and Reference of a Proper Name' p.160.
Dummett's point must be not that a modest theory presupposes an understanding of the language in which the meaning theory is stated but rather that it presupposes an understanding of the language in which the agent formulates the content of the theory.

Is this a valid criticism? It seems to me that it is not. For the whole issue of which concepts a person to whom a meaning theory for L is to be conveyed actually possesses seems to me to be a red herring - arriving at a theory of meaning for L oneself is one thing, communicating it to someone else is another.

Consider languageless Friday alone on a desert island before the advent of English-speaking Robinson Crusoe (to change the story a little). Friday has beliefs but no language pre-Crusoe; post-Crusoe he has a language and as a consequence an enormously expanded store of beliefs. In passing from one state to another by acquiring Crusoe's language, Friday, according to Davidson, has acquired a body of knowledge which can be analysed by a theory of meaning for English. Now let the meaning theory be as 'modest' as one likes so that it simply states the content of Friday's newly acquired knowledge without saying what that knowledge consists in or how Friday manifests it, then it could not sensibly be maintained that he surreptitiously relied upon his prior grasp of a background language in order to derive a knowledge of the object-language for ex hypothesi there is no such background language - Friday's grasp of English-as-a-metalanguage is coeval with his grasp of English-as-an-object-language. Changing the story once more, let Crusoe be a slightly deranged semantic theorist bent on testing the empirical adequacy of his meaning theory for English by seeing whether a languageless person can be taught the language by imbibing the contents of the theory. In this guise, Crusoe might well complain that if only Friday had
some language in which he could formulate the contents of the theory he
could learn to speak English.

I draw the following morals from these tales:
(i) Davidson’s project does not require that a radical interpreter already be in
possession of a language into which he can translate the foreign language,
although it does, naturally, require that the interpreter have the capacity to
speak a language. Interpretation is not translation and the supposed parallel that Dummett
draws between modest theories of meaning and translation manuals fails – Friday has no language
into which he can translate or in which he can formulate the contents of a meaning theory for
English.
(ii) Davidson’s theory is entirely neutral on the question of the conceptual
resources that the interpreter need be assumed to start with. Friday might have a
very spare store of concepts pre-Crusoe or a luxuriant stock of concepts inscribed in a language of
thought, for all the theory cares. Irrespective of whether Friday has a language of thought
precursor for every concept that English is capable of giving expression to or not, the theory of
meaning must show “which concepts are associated with which words”, i.e. give the meanings of the
expressions of English.
(iii) It is crucial to any attempt at evaluation of Davidson’s theory that we see a
Tarskian truth-theory for an object-language as being arrived at within a context
of radical interpretation. Only the first tale in which Friday (counterfactually) arrives at
the theory for himself and not the second in which a theorist subsequently attempts to
communicate a finished theory to him is of relevance to Davidson’s project.

The distinction between modest and full-blooded theories cannot then be
drawn in the way that Dummett first drew it at ‘What is a Theory of
Meaning?’ if it is to be of any relevance to Davidson’s programme.

Dummett contends that a full-blooded theory of meaning provides, whereas
modest theories fail to provide, a theory of sense. However, Dummett
places certain constraints upon acceptable theories of sense – most
importantly, that they should pair theorems of the truth theory (or ‘theory
of reference’) with certain specific practical abilities on the part of the
speaker. So, not every theory of meaning which includes a theory of sense

---

6 This is a rather tendentious charge because McDowell, an advocate of modest theories still claims
to be providing a theory of sense – cf ‘On the Sense and Reference of a Proper Name’. If we
understand by a theory of sense’ that part of a theory of meaning for a language which specifies
the contents of utterances of L-sentences, as McDowell does, then Dummett’s charge does seem
unwarranted. But Dummett is really after a theory of what the L-speaker knows qua speaker of L
and McDowell has consistently refused to answer this with anything other than a description of the
interpretative abilities of L-speakers – in this sense of ‘sense’ modest theories of meaning do not
provide a theory of sense (what in S2.2 I called a theory of sense).
will, in Dummett's eyes, pass as a *full-blooded* theory of meaning - the theory of sense must be of a definite kind.

McDowell has been a consistent advocate of modest theories of meaning. As we have previously seen, he conceives of a theory of meaning as bipartite in form, with a truth-theory specifying the contents of utterances and a theory of force specifying the type of speech act an utterance can be used to effect.

A modest theory of meaning refuses to pair practical abilities on the part of the speaker with known propositions of the theory of reference; practical abilities which can be perceived by any normal human observer - even one without a language - as the manifestations of *implicit knowledge* of those propositions. Indeed, the only response to Dummett's demand for manifestation that a modest theory will supply will be a 'theory-presupposing' one.7

Thus, a modest theorist will respond to Dummett's request to specify the behaviour of a speaker of French, say, which manifests his knowledge that 'Londres est jolie' means that London is beautiful by pointing to the practical ability of that speaker to use 'Londres est jolie' in such a way that fellow French speakers take him to be asserting that London is beautiful.

The meaning theory will record this 'implicit knowledge' in the form of a 'meaning theorem', or on the Davidsonian model a T-sentence - viz. "'Londres est jolie' is true ↔ London is beautiful." This is the distinctive piece of knowledge Pierre has of the meaning of 'Londres est jolie' by virtue of knowing French.

---

7 'On the Sense and Reference of a Proper Name' p.159.
§3.2.2 Dummett's objections to modest theories of meaning.

In 'What is a Theory of Meaning?', Dummett offers several arguments against the acceptability of modest theories of meaning. I believe that some of these arguments are unsound insofar as they trade on misunderstandings of Davidson's project; but there is one argument sketched there and developed in 'What is a Theory of Meaning (II)?' and most explicitly in The William James Lectures that is entirely correct - to wit, that modest theories fail to explicate the meanings of the logical constants and to justify thereby the logical laws assumed to hold in the metalanguage.

I shall begin with what I have alleged to be the unsound arguments in order to show that a Realist of Davidsonian persuasion need not be worried by them. One of these we have already seen to be flawed. Thus, Dummett's allegation that a modest theory of meaning:

'merely exhibits what it is to arrive at an interpretation of one language via an understanding of another, which is just what a translation manual does: it does not explain what it is to have a mastery of a language, say one's mother tongue, independently of a knowledge of any other.'

is, I believe, mistaken, not because it confuses use with mention or is impervious to the fact that translation manuals merely pair sentences with other sentences rather than with their contents, but because the case of languageless Friday reveals that there need not be any background language into which the theory is being surreptitiously translated. Moreover, that example establishes also that we can avoid that charge without attributing to Friday:

'a knowledge of the propositions expressed by the sentences of the theory of truth independently of any language in which those propositions might be expressed.'

---

8 "What is a Theory of Meaning?" p. 114.
9 loc. cit. italics mine.
Dummett offers an argument that purports to prove that Davidson can give no model of what a speaker’s knowledge of the theorems of a truth-theory for his language could consist in, i.e. that Davidson’s theory is modest\textsuperscript{10}. This argument makes much of a supposed distinction - that between knowing of a sentence that it is true and knowing the proposition expressed by the sentence. Dummett is at pains to establish a thesis that would seem quite uncontroversial - that an interpreter must know the contents of the clauses of the truth-theory he constructs for a speaker and not just know that those clauses are true. He apparently believes that Davidson must deny this uncontroversial thesis.\textsuperscript{11}

Why does Dummett think that Davidson is committed to denying that a speaker who knows a theory of meaning for his language knows the contents of its theorems? It is not entirely clear to me why he believes this but the following is my best attempt at understanding his argument:

\textsuperscript{10} Dummett later retracts this charge in the Appendix to “What is a Theory of Meaning?”, offering what I have elsewhere (§2.5) referred to as the ‘Stochastic Interpretation’ of Davidson - one on which the radical interpreter tries to maximise statistically the truth of the sentences the native holds true.

\textsuperscript{11} Perhaps Dummett’s distinction, or a weaker variant, might be of use in describing an example of Stephen Stich’s. Stich considers the (factual) case of a woman who in her youth had been distressed at the assassination of President McKinley, but in old age, afflicted by degenerative senility, could remember only that McKinley had been assassinated and was not even sure if assassinated people died (‘On the Ascription of Content’ p.191 in Woodfield, A (ed.) Thought and Object: Essays on Intentionality, Clarendon Press, Oxford, 1982, pp.155-206). Stich thinks that most people would not ascribe to the senile woman the belief that McKinley was assassinated, although we would not hesitate to ascribe this belief to her younger stage. Stich does not consider this but we might be able to ascribe to her the belief that “McKinley was assassinated” is true. I think we should hesitate to impute to her knowledge that the sentence was true.
(1) All that an interpreter interpreting a foreign language or a speaker acquiring his mother tongue have to go on, according to Davidson, are 'judgments of speakers as to the truth or falsity of sentences, and the conditions prevailing when these judgments [are] made.'

(2) Apropos the person R learning his mother tongue, he can know that sentences like "It is raining" is true iff it is raining" are true without having any inkling as to what 'It is raining' means, and he can know that sentences such as 'It is raining' are held true under certain conditions without knowing which conditions they are held true under.

(3) So how is R supposed to know what 'It is raining' means or what belief it can be used to express?

(4) If speakers invariably hold sentences true just when its truth-condition obtained, then N might be able to discern its meaning from their circumstantial behaviour - e.g. N's pointing to the sky and running for shelter whenever N uttered 'It is raining'.

(5) But this would only work in the general case if speakers were both infallible and possessed of superhuman epistemic capacities. For since speakers make mistakes, R cannot just read off the truth-conditions of their sentences whenever those truth-conditions in fact obtain, and since speakers are incapable of recognising the truth-conditions of many of their sentences even when they do obtain, this model is only of use for a very restricted class of sentences - ones that their actual epistemic capacities can cover.

(6) Davidson, whilst claiming that R need only know that the axioms and theorems of the truth-theory are true, will try to avoid the conclusion that R will not be able to discern the truth-conditions of sentences by claiming that the truth-theory must be construed holistically - R need not know the proposition expressed by "London denotes London", he need only know that that axiom is true and that it is an axiom of a truth-theory that achieves the closest fit, given the fact that speakers do make mistakes, with the sentences those speakers hold true.

(7) But this holistic response will not work for three reasons:
(i) Holism cannot allow for mistakes since it cannot distinguish disagreements about meaning from disagreements about the facts.
(ii) What tells R what 'London' means is the axiom "London denotes London" and not any information about the axiom.
(iii) Although it purports to provide a theoretical representation of the whole practical ability to speak the language through a deductively connected set of propositions, the appearance of structure is a fraud just because it gives no model of what knowledge of any individual proposition consists in.

12 "What is a Theory of Meaning?" p.120.
This is a very dense argument and I shall not try to analyse it in its entirety. However, I think I can show that the crucial second premise at least is false.\(^{13}\)

Premise (2) represents my attempt to find a place for the distinction that Dummett makes so much of in his argument against Davidson. It is obviously correct that the mere knowledge that a certain sentence possesses a certain semantic property (viz. truth) neither suffices for nor requires knowledge of what the sentence means. Perhaps the radical interpreter R can know (as a purely syntactic fact about English) that ""It is raining" is true iff it is raining" is true. Why should this be of interest to Davidson? The T-sentences of a truth-theory represent empirical generalisations, according to Davidson, that the interpreter has hit upon himself. How could R, after having laboriously tested them against the behavioural evidence provided by speakers, fail to understand those very T-sentences he has himself arrived at? Dummett seems to have conflated the genuine question of how a truth-theory can be constructed from the evidence with the irrelevant issue of how it could subsequently be communicated to one who was not involved in the construction of that theory. R does not, on Davidson's picture at least, learn of the truth of various T-theorems second-hand, through the testimony of a field linguist who is actively engaged in interpreting the speakers in question. Why, then does Dummett misconstrue Davidson thus?

\(^{13}\) I suppose the first premise is false as well, since Davidson never held that knowledge of the sentences N holds true is the only evidence R can reliably ascertain from N — he can also find out which sentences N prefers true or intends to make true etc. Indeed in recent writings Davidson has made preferring-true the cornerstone of his programme (see 'A New Basis for Decision Theory' Theory and Decision, 17, 1985).
The answer, I think, lies in premise (6). Dummett thinks that Davidson's replacement of Tarski's Convention (T) demand (that p translate s in the schema: 's is true iff p') by the indirect holistic constraint: 'That the totality of T-sentences should ... optimally fit evidence about sentences held true by native speakers.'

is tantamount to an admission that the interpreter cannot be assumed to know the proposition expressed by any T-sentence, he can only know that it is true.

But this is not what Davidson means at all. What the interpreter cannot assume is that a T-sentence that looks initially plausible - e.g. (A) "'It is raining' is true iff the speaker cannot find his overcoat" - is an interpretational T-sentence without checking that it is a T-sentence derivable in a formally and empirically adequate (i.e. interpretational) truth-theory. For this would be to assume that he could know from the outset what it is the object of radical interpretation to find out - that the RHS's of putative T-sentences translated the LHS's. But there is nothing in this to even suggest that N must therefore only know that (A) is false without knowing what it means or that the correct T-sentence for 'It is raining' is true without knowing what it means; indeed it would reduce Davidson's method to absurdity if this were to be so since sentences such as (A) are supposed to represent the interpreter's own conjectures.

There is therefore no substance to Dummett's charge that a radical interpreter cannot know the propositions expressed by the clauses of the truth-theory.

But there is a separate question Dummett raises which Davidson cannot duck so easily - on what grounds could we attribute knowledge of the propositions

14 'Radical Interpretation' loc. cit. p. 139.
of the truth-theory to a speaker-interpreter? When the truth-theory assumes that classical logical laws hold in its proof theory, this is equivalent to the Anti-Realist challenge to the Realist: what in a speaker's behaviour could possibly warrant ascribing to him a knowledge that \( s_k \text{ is true} \) iff certain conditions which are possibly verification-transcendent obtain?

Dummett thinks that Davidson will want to reply (cf premise (6)) that whilst there is no provision on his theory for a model of what a knowledge of the individual propositions of the truth-theory consists in, the whole theory is to be understood as modelling the entire practical ability to speak the language.

Dummett replies (cf 7(iii)) that this response will not do because:

'A speaker's knowledge of the meaning of an individual sentence is represented as consisting in his grasp of a part of a deductive theory, and this is connected with his actual utterances only by the fact that a grasp of the whole theory is supposed to issue, in some manner of which no explanation is given, in his command of the language in its entirety; but no way is provided even in principle, of segmenting his ability to use the language as a whole into distinct component abilities which manifest his understanding of individual words, sentences, or types of sentence. To effect any such segmentation, it would be necessary to give a detailed account of the practical ability in which the understanding of a particular word or sentence consisted, whereas, on the holistic view, not only cannot a speaker's command of his language be so segmented, but no detailed description of what it consists in can be given at all. Hence the articulation of the theory plays no genuine role in the account of what constitutes a speaker's mastery of his language.'

Dummett's general complaint against modest theories of meaning is therefore that they provide no model of what a speaker's knowledge of the individual propositions of the truth-theory might consist in.

This is most in evidence when we ask what constitutes an understanding of the logical constants of the object-language on a modest theory of meaning. Suppose, as on the 'modest' theorist's favoured model, that we are dealing with a homophonic truth-theory for English - one in which the metalanguage is an extension of the object-language. It is surely within the realms of possibility that a certain English-speaking community might understand the

\[15\] 'What is a Theory of Meaning?' p. 116.
logical constants intuitionistically rather than classically. Dummett thinks there is a problem in how an interpreter, or a speaker learning his mother tongue, can allow for this possibility if the meaning theory he constructs is a modest one.

A homophonic truth-theory contains 'disquotational' clauses for atomic and complex sentences of the form ""p" is true iff p' and ""p or q" is true iff p is true or q is true'. Dummett calls such clauses straightforward stipulations at The William James Lectures since, in the case of the clauses for compound sentences, the truth of the complex is stated straightforwardly in terms of the truth of its constituent sentences.

Now a truth-definition for a language can be constructed in a metalanguage which uses a non-classical proof-theory provided only that the truth-predicate distributes over the logical constants. Such a metalanguage will read into the object-language a non-classical interpretation of its logical operators. Similarly, if the proof-theory is classical, the object-language will be ascribed a classical logic. In either case, the straightforward stipulations will be the same; and it will depend entirely upon the laws assumed to hold in the metalanguage whether the object-language logic is taken to be classical or non-classical. Yet it seems an unwarranted assumption that object-language logic and metalanguage logic coincide.

Although he does not regard a homophonic semantics as useless, Dummett thinks that it cannot be used at even the most cursory level to make explicit the senses of the logical constants since it perforce assumes that the constants are being used with their intended meanings. Moreover, when we try to construct a semantic theory which yields straightforward stipulations

---

16 He calls them trivial axioms at "What is a Theory of Meaning (II)?", but I shall use the William James Lectures terminology as I think the use of 'trivial' is a little tendentious.
17 Cf 'What is a Theory of Meaning (II)?' p.107.
for the relevant complex sentences, we find that the resultant theory is quite uninformative in explicating the senses of the constants wherever those constants are non-classical.

It is the goal of every semantic theory to show how the truth of a complex sentence is determined in accord with its structure. Classical bivalent semantics, however, is alone in taking the semantic value of a sentence to be its truth-value - for non-classical semantic theories, the truth-value of a compound sentence will not just depend upon the truth-values of its constituent sentences. So, whilst it is always possible to state the truth of a complex sentence *in terms of* the truth of its sentential parts, it is only for the special case of classical semantics that its truth is *determined by* that of its sub-sentences. Since intuitionistic semantics does not take the semantic value of a sentence to be its truth-value, the formulation of the truth of a complex sentence in terms of the truth of its sub-sentences does not reveal how the semantic value of the complex depends upon that of its parts, nor, therefore, how the truth of the complex depends upon its semantic structure. Dummett calls an interpretation which proceeds by stipulating straightforwardly the truth of a complex in terms of its parts, an *Internal Interpretation*. Because internal interpretations do not meet these most fundamental of goals for semantic theories, Dummett doubts whether we should think of them as genuinely semantic interpretations at all.

Moreover, we can verify this suspicion by seeing what the semantic value of a sentence is taken to be on an intuitionistic Internal Interpretation - the answer is: a proposition. Nothing in the Internal semantic theory however can tell us what a proposition is understood to be, it is just assumed that we know. Internal Interpretations are therefore wholly *Programmatic Interpretations*. 
There is an intuitionistic semantics that casts light on the question of how the intuitionist thinks of propositions - Heyting's theory of constructions. On that theory, a proposition is a decidable classification of constructions into those that do and those that do not prove the sentence which conveys it. But, since Heyting's stipulations are not straightforward, we cannot incorporate his explanations of the logical constants into an Internal Interpretation. Since there is no standard semantics for intuitionism and we must therefore specify the concept of validity we are working with, we have no option other than to abandon straightforward stipulations if we wish to give more than a Programmatic interpretation of the semantic value of a sentence.
§3.3 OTHER MINDS

§3.3.1 The Problem for Anti-Realism

The Anti-Realist’s challenge to the Realist is to explain how an understanding of a verification-transcendent state of affairs could be acquired or manifested. Strawson has replied on behalf of the Realist that there are two domains for which we do in fact possess such a conception – that of the past and of other minds.1 Dummett and Wright have not been persuaded by Strawson’s claim2. Dummett has concentrated upon the problem of the past3 whilst Wright has sought to explicate our grasp of statements about the mental states of others without appealing to any verification-transcendent notions4.

Strawson had claimed:

"It is part of what it is now regrettably fashionable to call our general theory of the world that we regard other people as subject to roughly the same range of sensations as we are... And it is in no way contrary to reason to regard ourselves, as in any case we cannot help doing, as justified in certain circumstances in ascribing to John a particular state of feeling which we cannot in the

---

1 'Scruton and Wright on Anti-Realism etc.' loc. cit.
2 Dummett: Truth and Other Enigmas preface xxxii – xxxviii; Wright: 'Strawson on Anti-Realism.', loc. cit.
3 See Essays 19, 20, 21 in Truth and Other Enigmas. Dummett thinks that statements about the past present far more of a challenge to Anti-Realism than statements about other minds [TOE xxxviii]. He criticises Strawson for believing that we know what it is for another to be in pain because we have given ourselves a private ostensive definition [ibid xxxii] and records his conviction that Wittgenstein’s polemic against the possibility of such a procedure is “incontrovertible”. I am not convinced that a Realist about other minds such as Strawson need believe in ‘private’ ostensive definitions, so I am not sure that Dummett’s reply is to the point. Be that as it may, Dummett reveals that he has not seen the real worry for the Anti-Realist’s position for he takes it to be about grasping the truth-conditions of statements like ‘John is in pain’ (cf: “To understand statements like ‘John is in pain’, we must know how they are used “[ibid xxxv]”)

I hope to show that it is not these statements which are particularly problematic for the Anti-Realist.

4 See the articles cited in §2.3: ‘Strawson on Anti-Realism’; ‘Realism, Truth-Value links, Other Minds and the Past’; Anti-Realist Semantics: the role of Criteria’; ‘Second Thoughts about Criteria’.
very nature of the case experience ourselves and his being in which is therefore
... necessarily verification-transcendent. 5

Commenting upon this passage, Crispin Wright wrote:
'The assumption is that to oppose Realism is to call into question these aspects of our 'general
theory of the world'. But it is open to an Anti-Realist to seek to interpret them. Thus, the
susceptibility of others to the same range of sensations as we ourselves may be viewed as an
expression of the existence of a communal vocabulary of sensation, any element of which is
applicable to any of us on the basis of communally acknowledged criteria and the inaccessible of
others' states of feeling may be viewed as an expression of the essential defeasibility of other-
ascriptions of sensation, the fact that any state of information which warrants such an ascription
can always coherently be envisaged as being added to in such a way that the resulting state of
information no longer does so. By contrast, the 'privileged access' we are traditionally thought to
have to our own sensations may be viewed as an expression of the fact that our grammar of
sensation provides for the possibility of discovering that one did not know what one was saying in
making a particular self-ascription, but not for the possibility that one understood it and was
mistaken. 6

It seems to me that both Wright and Dummett have simply failed to
appreciate the particular problem for Anti-Realism posed by other minds.
Ironically, it was the later Wittgenstein whose views both Dummett 7 and
Wright 8 take themselves to be closely approximating who best articulated
the problem Strawson raises:
'If one has to imagine someone else's pain on the model of one's own, this is none too easy a thing to
do for I have to imagine pain which I do not feel on the model of the pain which I do feel... For
I am not to imagine that I feel pain in some region of his body. 9

Someone who utters 'I am in pain' will, if sincere, characteristically be
experiencing some painful sensation but not, or not usually, be perceiving a
person writhing or grimacing. An onlooker reporting this event as 'Jones is in
pain', sees a person, whom he identifies as Jones, writhing or grimacing, but
he does not experience his sensation. Truistic though these facts be, they are

5 loc cit.
6 'Strawson on Anti-Realism' loc. cit. pp. 289-290. Wright then comments "naturally, this is only
the sketch of an approach, and I cannot claim to know that it would prove satisfactory in detail".What would prevent such a conventionalist account from succeeding? Whatever facts about
sensations are produced, the conventionalist will simply 'account' for them as unremarkable facts
about the language-game of sensation.
7 loc. cit. pp.xxii ff.
sufficient to generate a problem which is particularly vexing for the Anti-
Realist.

Let us call the mode of epistemic access Jones has to his own pain an
'internal' mode of presentation (m.o.p) of pain and the mode of access others
apart from Jones have to Jones' pain, an 'external' mode of presentation of
pain. An internal m.o.p of pain derives from an experience of pain; an
external m.o.p of pain derives from the observation of the behavioural
manifestations of pain. The problem in the 'problem of other minds'
derives from the fact that there is an epistemic asymmetry between Jones
and everyone else with respect to Jones' pain - Jones has both an internal
and an external m.o.p of his own pain, everyone else has only an external
m.o.p. The conditions for justifiably asserting 'I am in pain' in the case of
Jones and 'Jones is in pain' in the case of anyone else, differ accordingly.

There now arises a problem for the Anti-Realist:

If the meaning of a statement is given by its canonical assertibility
conditions, the Anti-Realist must show how 'I am in pain' uttered by Jones
and 'Jones is in pain' uttered by me could possibly mean the same or, at the
very least, report the same state of affairs. The canonical assertibility
condition for the former is that Jones experience, or at the very least believe
himself to be experiencing, a pain; the canonical assertibility condition for
the latter, that the person satisfy the criteria of identity for Jones and that
he evince a reliable behavioural manifestation of pain - e.g. that he wince or

10 I am making no presumption that the internal m.o.p is incorrigible, only that it is available to
the subject of experience alone.

11 It is a little hard to know whether to put this point in terms of sameness of meaning, given that
the Anti-Realist need not be committed to propositions. Still, we can formulate the point in terms
of reporting the same state of affairs or event.
moan or sincerely claim that he is in pain\textsuperscript{12}. Since canonical assertibility conditions differ, the meanings of the two statements must also differ. Two problems thus face the Anti-Realist. He must show that the predicate 'pain' has a unitary sense even though its canonical criteria of application differ so starkly in first and third person exemplifications. Anti-Realistically, the presumption must surely be that 'pain' in these separate exemplifications differs in sense - let us call the internally presented instance \textit{pain} and the externally presented instance \textit{pain*}, the former a sensation, the latter a cluster of behavioural criteria. I shall call it the problem of Semantic Variance that 'pain' is equivocal as between pain and pain*.

Secondly, he must allay the suspicion that it is belief in the verification-transcendent state of affairs of Jones' pain which ultimately warrants the assertion of 'Jones is in pain'.

How then should the Anti-Realist best respond to the problems above? Semantic Variance suggests that when Jones says 'I am in pain', by 'pain' he means \textit{pain} and that when I report this as 'Jones is in pain', by 'pain' I mean \textit{pain*}. Trivially, I cannot then be reporting the same event as Jones is reporting. Unless he is to somehow show that, contrary to all appearances, it is innocuous, the Anti-Realist must somehow block Semantic Variance\textsuperscript{13}.

\textsuperscript{12} Can the Anti-Realist accept sincere avowals about conscious experiences not only as a criterion that the subject has those experiences but as the ultimately decisive criterion? This seems to me to place too much faith in the understanding that the agent has of his language, particularly if the view propounded is a Wittgensteinian one in which 'pain' does not refer to the sensation of pain.

\textsuperscript{13} Dummett clearly rejects it: 'It is obviously correct, independently of any philosophical position, to say that 'John is in pain' is true when and only when it is with John as it is with me when I am in pain.' loc. cit. T.O.E. p.xxxiv
The obvious move is to deny that Jones has an internal m.o.p available only to him. We were wrong in thinking that 'pain' was the name of a certain sensation. This seems to have been Wittgenstein's position.

Let us index the 'I' in Jones' utterance 'I am in pain' as 'I_{Jones}' or just 'I'. Then we wish to know how to assign assertibility conditions to 'I am in pain'.

Wittgenstein's position looks like Behaviourism. Dummett claims that it is not - that both Behaviourism and Realism only arise when we illicitly assume that our understanding of pain-ascriptions is a truth-conditional one.

The thrust of the argument, Dummett tells us, is first, that we cannot employ the notion of reference to explain how expressions for inner sensations function ('if we construe the grammar of the expression of sensation on the model of "object and name", the object drops out of consideration as irrelevant' - Philosophical Investigations 5293) and, secondly, that an understanding of ascriptions of inner sensation cannot be explained as consisting in a knowledge of the condition for them to be true.

Insofar as I grasp this Wittgensteinian point at all, it seems to be that the truth of 'x is in pain' does not involve the identification of any object: viz. the referent of 'pain'. To the question "what does 'pain' refer to?", we can only give a trivial reply: "'pain' refers to pain", just as we can only provide a trivial axiom in response to the request for the truth-conditions of 'Jones is in pain' - viz, "'Jones is in pain' is true iff Jones is in pain". So, it is wrong to expect an understanding of 'x is in pain' to be obtained by knowing its satisfaction conditions. How then do we come to understand it, according to Dummett?

'To understand statements like 'John is in pain', we must know how they are used. That involves knowing that pain-behaviour, or the presence of an ordinarily painful stimulus, is normally a sufficient ground for an ascription of pain, but one that can be rebutted, in the former case by the clues that betray the shammer or by subsequent disclaimer; learning the symptoms of inhibiting the natural manifestation of pain, and the limits beyond which this is impossible; knowing the usual connection between pain and bodily conditions, and the sort of cases in which the connection may be broken, and so on. To know these and similar things, on Wittgenstein's account, just is to know what 'John is in pain' means; and, for one who knows this, there need be no more informative

14 loc. cit. pp.xxxiii-xxxv.
answer to the question what makes that statement true than, 'John's being in pain'. We looked for an informative answer, midway between those of the realist and the behaviourist, when no informative answer was to be had.\footnote{loc. cit. pp.xxxv.}

There are several things to note about this passage. First and foremost, the account of the assertibility conditions of 'John is in pain' is simply incoherent unless 'pain' is taken to refer to the sensation of pain. To see this, suppose we try to use the account to inform a Putnamesque superSpartan who successfully inhibits the behavioural manifestation of his sensations, moods and emotions what the word 'pain' means. When the superSpartan is in pain, the red light on his C-fibre monitor lights up, when he is afraid, the D-fibre monitor lights, when happy, the E-fibre monitor lights.

Now Dummett's account works just as well as an account of the assertibility conditions of 'John is afraid' or 'John is happy'. How is our friend to know whether 'pain' means happy or afraid rather than pain? Until he advances the hypothesis that 'pain' refers to a sensation similar to the one he experiences when his C-fibre monitor lights, he will surely be completely lost both as to the meaning of 'pain' and as to the explanation of the motley of wincings, curses, moans, etc he observes in connection with its use. Contrary to Dummett then, we must employ the notion of reference, to explain how sensation terms function, as his own theory demonstrates.\footnote{Dummett will most probably object that there has been an illicit Realism appealed to in the description of the case but it is not clear to me that there need be any conceptual incoherence from the Anti-Realist's point of view in the possibility that a creature could always inhibit certain of its natural behavioural dispositions if it had enough motivational strength. Perhaps the superSpartan thinks he will be struck dead by Zeus and eternally damned if he so much as murmurs when he loses a limb. That he should believe this seems to present us with a recognisable possibility.}

Secondly, given that 'pain' does refer to the sensation of pain as it must for the above account to be coherent, Dummett's explanation of the assertibility conditions of 'Jones is in pain' is precisely the type of explanation a \textit{Realist}
would give of that statement’s assertibility conditions. Realism is emphatically not the (absurd) thesis that a grasp of assertibility conditions plays no part in a model of a speaker’s competence with a given range of sentences — on the contrary, it will be crucial to an understanding of the meaning of ‘Jones is in pain’ that a speaker knows what would constitute good grounds for believing it to be true. Since for a Realist, a sentence is assertible just when there are good grounds for believing it to be true, not to know what comprised such grounds would result in a failure to know when one could legitimately assert the sentence.18

Doubts about the internal coherence of Dummett’s theory aside, how does it deal with the problem of Semantic Variance? It doesn’t: it simply ignores it. We are left completely in the dark as to what the assertibility conditions of ‘I am in pain’ are19 and as to whether these are the same as ‘Jones is in pain’ or not. The fact that ‘pain’ is not the name of a sensation seems to prevent ‘I am in pain’ being assertible just when I experience a certain sensation, although even this is unclear. If so, how are we to communicate to superSpartan Jones the assertibility conditions of ‘I am in pain’? Nothing observable is of any use here unless one accepts a Behaviourist reduction of the meaning of the statement.

18 Dummett apparently does think that the Realist, in this instance at least, must deny that a grasp of the assertibility conditions of ‘John is in pain’ plays any part in understanding the statement, for he says at p. xxxii of the Preface that on the Realist view: ‘... the sense of the sentence is not given in terms of the evidence that I can have for its truth or falsity: such evidence is necessarily indirect and often inconclusive; but my knowledge of what the sentence means is independent of my knowledge of what constitutes such evidence’. But perhaps Dummett means to be making a slightly different point: that the Realist holds that grasp of the assertibility conditions of the sentence is somewhat contingent upon a prior grasp of its truth-conditions. This certainly does present the naive Realist who believes that we understand the sentence by pairing it off with a verification-transcendent state of affairs, with a severe difficulty. I discuss this in more depth in §3.4.

19 Or for that matter what those of ‘I am in pain’ are for any i.
One suggestion, consistent with Dummett's theory and again inspired by some of Wittgenstein's remarks, is that 'I am in pain' is simply a more sophisticated way of saying 'Ouch!'. The sentence expresses pain, it does not describe it. It is a semantically unstructured or 'holophrastic' phrase (a la Quine's observation sentences). Then, 'I am in pain' and 'Jones is in pain' do not have the same assertibility conditions; we are thus to acquiesce in Semantic Variance. Call this the Holophrastic Thesis. Then this seems to be a straightforwardly Behaviouristic thesis and thus to suffer from all the problems Behaviourism suffers from. If accepted, it would mean that our mental language would have to be extensively revised. Existential Generalisation and Universal Instantiation would no longer apply to 'I am in pain'. This has some rather bizarre consequences:

Suppose amnesic Jones lies groaning in agony in a hospital bed. A doctor breezing by remarks to the nurse who accompanies him "There is one and only one person in pain in this ward"; "Yes", the nurse responds, "Jones is in pain"; Jones upon hearing this snatch of conversation says out loud "Well I am in pain, so I must be Jones!". If the Holophrastic Thesis were correct, it would follow that Jones is mistaken in inferring thus; that, supposing the doctor's and nurse's remarks to be true, he would have acquired knowledge of who he was by illicit means; that the doctor and the nurse could not have conspired to truly inform him of his identity in this way. For the meaning of 'pain' differs in Jones' mouth from its meaning in the mouths of the doctor and the nurse.

The Anti-Realist might seek to avoid the problem of Semantic Variance by denying that pain and pain² are independent properties - it is just that one's own experience of pain includes an element incommunicable to any one
other than the sufferer: the pain quale. But since it also includes communicable manifestations - behavioural or verbal - there is no equivocation: the states of affairs under which 'I am in pain' is assertible comprehend just those states of affairs in which 'Jones is in pain' is assertible. So whilst 'I am in pain' assuredly has a richer cognitive content than 'Jones is in pain' for Jones, the semantic information conveyed in both cases to an onlooker is identical.

This position would be tenable only if there were entailments between 'I am in pain' and behavioural manifestations of some sort, but there need not be any such manifestations, outside of the verbal report at all - we cannot accept the verbal report itself as such evidence unless we know what it means, and the question precisely is what the assertibility condition of the statement, hence its meaning, is. Moreover, even if we were to grant that, per impossibile, non-verbal behavioural manifestations could in principle be observed for any utterance of 'I am in pain', it still would not follow that the circumstances in which this statement can justifiably be uttered invariably comprehended the circumstances in which 'Jones is in pain' could be so asserted, at least so far as Jones is concerned. For consider amnesic Jones again and imagine that the doctor is a deranged sadist who has completely refashioned Jones' body whilst he was comatose (perhaps transformed it into a body that is externally that of a woman's), then non-comatose Jones, who asserts 'I am in pain', will, upon learning once more of his name and seeing his own woman's body contorted in pain, deny 'Jones is in pain', on the evidence of the writhing body he sees. Even after the grisly truth is revealed to him, it still might remain that Jones is never inclined to assert, in point of fact he might
be still inclined to deny, 'I am in pain' on the basis of an external m.o.p available to others.

As far as I can see, the Anti-Realist simply has to admit the existence of Semantic Variance between 'I am in pain' and 'Jones is in pain'. The thing to do then is to try to play down its importance. It would be quite unconvincing to attempt to argue that the class of mental phenomena which have internal m.o.p's is too small to worry about - it is surely no smaller than the whole range of moods, feelings, sensations etc. which agents consciously experience. He could contend along Paul Churchland's lines that the language we use to report not only our sensations but also our conscious experiences at large is just incoherent at heart - that our scientifically astute successors would revise on neurophysiological grounds our folk-experiential vocabulary: then both 'I am in pain' uttered by Jones and 'Jones is in pain' uttered by anyone else, would be assertible just when the red light on Jones' C-fibre monitor lit up. Until such times as neurophysiology develops, we are left with semantic paradoxes in our attempts to systematise the assertibility conditions of psychological statements.

There may be other, more plausible responses an Anti-Realist can make to the problem of Semantic Variance but I cannot see how he can avoid it. So the question still remains: 'How is the Anti-Realist to assign assertibility conditions to the first person psychological utterances of others?'. My conjecture is that Anti-Realism cannot provide any plausible model of the

---

20 This raises the question of whether Jones' external m.o.p is the same as the external m.o.p available to others. The claim that it is seems congenial to Cartesianism.

21 Scientific Realism and the Plasticity of Mind ch. 1.

22 Whether Anti-Realism is compatible with Eliminative Materialism is a question I shall not pursue.

23 In order to avoid it, the Anti-Realist must deny that Jones' utterance of 'I am in pain' has as its canonical assertibility condition that Jones experience or believe himself to be experiencing, a pain.
conscious experiences of others. But at this stage, we do not know whether Realism can either.

§3.3.2 Realism and the conscious experience of others

The Realist is, prima facie, untroubled by Semantic Variance - 'I am in pain' and 'Jones is in pain' have, he holds, a common truth-condition: Jones' being in pain; a state of affairs which is, in Strawson's words, 'necessarily verification-transcendent' to all but Jones. Can the Realist answer the Anti-Realist challenge to say how a conception of such a state of affairs is manifested in behaviour?

Recall Wittgenstein's challenge: to imagine the pain of another - pain which I do not feel - on the model of my own pain which I do feel, without mistakenly assimilating that to imagining my pain in another body. Colin McGinn, in discussing Wittgenstein's challenge, suggests that if we but had a m.o.p of pain which was neutral as between first person ('internal') and third person ('external') perspectives, we could be assured that our concept of pain in its first and third person exemplifications was unitary. The problem, he contends, is that no 'perspective-neutral' conception of pain is

24 'Consciousness and Other Minds' p.11.
available. I agree that we have no such m.o.p, but I do not know that it is required in order to answer Wittgenstein's challenge25.

Can one imagine the pain of another on the model of one's own? One can certainly imagine, pace Berkeley, an unperceived tree, say, on the model of a perceived tree. If one can imagine an unseen tree, why not an unfelt pain? if Wittgenstein means to be challenging us to imagine precisely how Jones' pain must feel to Jones in the sense of conjuring a sensuous image qualitatively identical to one that Jones would conjure, then we must agree that this cannot be done. But it does not seem as if this was what he meant. On analogy with visualisation by means of visual images, let us call the activity of imagining to feel a sensation via a sensuous image, 'sensuisation'. Now just as it might be held, plausibly, that visualisation consists in thinking of oneself seeing, so it might be held that sensuisation consists in thinking of oneself feeling, and then again it would not be possible to imagine Jones' pain by sensuising it. But Bernard Williams has argued very persuasively that we ought not to think of visualisation in this way26. Although visualisation of an object must involve visualising it as from some visual perspective or other, this does not entail that any imagined seeing is going

25 Someone might be tempted to try to answer Wittgenstein from the resolutely 'third-person' perspective of functionalism. We can dispose of his worry if we identify pain with a certain functional state would be the claim. Then we would have a univocal predicate 'x is in pain' which takes me and Jones as instances and thus shown that Jones and I can be in the same mental state by virtue of being in the same functional state. But this does not show that 'I am in pain' and 'Jones is in pain' express the same proposition. The Functionalist has only shown that 'I am in pain' uttered by me ascribes the same state to me as 'Jones is in pain' ascribes to Jones: viz pain. He has not shown that in ascribing a (functional) state of pain to Jones I am ascribing the very same thing Jones ascribes to himself. Wittgenstein's problem is thus untouched since I am ascribing pain that I do not feel to Jones, if not on the model of pain which I do feel, then through a theory that is, arguably, only intelligible on the model of my own pain.

on within the visualised scene. For it is not necessary that I belong inside the world I visualise any more than I need belong inside the world of the stage when I see, for example, Lear putting out his own eyes. There is no barrier then to our visualising an unseen object and no reason to think that the visual focus in visualisation must be mine.

27 loc. cit. p.34.
28 loc. cit. p.37.
But surely sensuisation differs just here, for it is absurd to think that there might be sensuisation of a feeling without someone's sensuising that feeling. Certainly. But why should we think that the sensuous focus must be mine? I might be able, within my imagination, to do the sensuising without occurring as the person who sensuises the pain; and I clearly cannot be that person if I am to sensuise Jones' pain. It would therefore seem that in order to fulfil Wittgenstein's requirements, I must imagine being Jones, and as Jones, being in pain. Imagining being Jones is not imagining my ordinary empirical self to be stripped of its own identity and then reconstituted (somehow) as Jones, Williams argues, for this is logically impossible. What I must do to fulfil that part of my imaginative project is, to act out the role of Jones in my imagination as I would act out the role of Lear. Then, in that imaginative mode, the 'I' as it occurs in my anguished cry 'I am in pain', refers to Jones if it refers at all. I could then imagine Jones' pain by sensuising pain on the model of my own, and the sensuous focus in my imagined setting would be Jones' and not mine. It will be objected: 'But you do not know what it is like to be Jones, how it is for him - for all you know, he might experience a feeling of euphoria whenever he treads upon red-hot coals and call that 'pain', so how can you imagine his pain?'

In this strong sense of 'what it is like to be Jones' or 'how it is for Jones', one could only know this if one were Jones, and I have agreed that this is impossible. But why should that present an obstacle to the more modest project of imagining the sensuous character of his pain? It seems possible\textsuperscript{29} that the pain that I imagine in my imaginative enactment might sensuise the pain Jones actually experiences at \( t_0 \) just as accurately as Jones' own

\textsuperscript{29} Such a possibility is of course verification-transcendent.
subsequent sensuisations at $t_n$; if so, why is there more of a problem in my case than there is for Jones$_n$ to imagine the pain of Jones$_0$?

If this line of thought is on the right track, we would have succeeded in imagining another's pain on the model of one's own without attempting, incoherently, to avail ourselves of a m.o.p not open to us - e.g. Jones' own, or worse, one neutral as between Jones' internal and any other's external m.o.p. There would be no pretence that we have somehow solved the problem of other minds or gained ingress into a verification-transcendent state of affairs, since the imaginative excursus respects the verification-transcendent status of Jones' pain. But can a Realist know that what seems as if it might be a possibility really is so, is his conception of Jones' pain really manifestable in his behaviour?

A Realist might be tempted to reply along the following lines here:-

'The conscious states of others pose precisely the same conceptual difficulties as those posed by any other theoretical entity. The justification for believing in pains, beliefs, and moods is precisely the same as the justification for believing in atoms, gravitons or black holes. In both cases the hypothetico-deductive method is at work. By hypothesing that Jones enjoys the same range of conscious experiences as we do, the seamless web of his behaviour reveals certain patterns - his crying out, grimacing and pain-avoidance behaviour become explicable as effects of a common cause: an unpleasant sensation. We find that he will behave as we predict that he will behave - he avoids losing limbs and walking on hot coals with all the commitment that we do. By conjoining the hypothesis that Jones suffers pain like us to the evidence of his behaviour and the observable circumstances we can infer consequences about the rest of his behaviour; consequences which are borne
out again and again by what Jones subsequently does. It is the explanatory
capacity and predictive success of the network of principles and
generalisations that correlate mental states with behavioural, environmental
and bodily circumstances and with further mental states which justifies belief in the hypothesis that Jones enjoys both sentience and sapience.\(^{30}\)

In order to evaluate this thesis, we must decide what is meant by 'behaviour'. If by 'behaviour' we simply mean bodily movements, extensionally described, then it is not at all obvious that the postulation of mental states is ultimately required for the task of explaining Jones' behaviour even if it is heuristically indispensable for predicting that behaviour. Laplacian-minded Martians would, presumably, be far more successful in predicting at the neurophysiological level what Jones would do next and why. The 'Intentional stance' would then have no abiding significance \textit{sub specie aeternitatis}, its theoretical entities 'abstracta' - calculation-bound entities rather than 'illata' - posited theoretical entities\(^{31}\).

If, on the other hand, 'behaviour' means actions intensionally described then our reason-giving descriptions do presuppose the existence of mental

\(^{30}\) It is crucial that this thesis be thought of as an answer to the question of how one knows that others have minds. It has no plausibility at all for one's own case since there one simply experiences pain, one does not infer to its existence from observations of its effects upon one's own behaviour except perhaps in highly unusual circumstances. Paul Churchland defends the thesis above at \textit{Matter and Consciousness} , Bradford Books, M.I.T. Press, Cambridge, Massachusetts, 1984, pp. 70-72; however he goes on to argue for its application to one's own case at pp. 73-81.

But surely the problem of other minds could not arise if there was this symmetry. Churchland confuses the question of whether an agent has an internal m.o.p available only to him with the question of whether that m.o.p is incorrigible.

\(^{31}\) As Dennett expresses it: 'What it is to be a true believer is to be an intentional system, a system whose behaviour is reliably and voluminously predictable via the intentional strategy.' 'True Believers: The Intentional Strategy and why it works' in \textit{Heath, A (ed.) Scientific Explanation}, Oxford University Press, Oxford, 1981, p.55. This general thesis seems strongest for the propositional attitudes, it seems highly unlikely that a developed neurophysiology would displace the felt character of the experience of pain with something else as the genuinely causally efficacious property etched out by 'pain'. See §1.1.
states of a very particular sort, and the continued explanatory success of the 'Intentional Strategy' provides good reason for believing that those states are illata rather than abstracta\textsuperscript{32}. However, it seems very unconvincing to maintain that we know that others have a mental life \textit{at all} by such reasoning\textsuperscript{33}. For if it is just a hypothesis that they do, it must be possible that this hypothesis is false. Yet the assumption that they do not exist leads to absurdity - not only can we not make sense of the world if we deny or even doubt that others have minds, we cannot make sense of our own conscious experiences\textsuperscript{34} (and not just 'other-implicating' experiences such as shame or love or envy).

How could a speaker manifest a grasp of a verification-transcendent truth-condition for \textit{'I am in pain'} on the Realist's account? The Realist's answer is that a speaker will manifest his grasp of the classical truth-condition of \textit{'I am in pain'} by his evaluations of the evidence for its truth and by the arguments featuring that sentence as assumption, premise or conclusion that he accepts as valid. If he sees Jones silently smirking as he says it, this will count as evidence against its truth; if he has some medical knowledge he may know how to test its truth via physiological symptoms typically evinced.

\textsuperscript{32} Though Dennett of course disagrees. Whether these mental states will themselves be required by a more explanatory psychological theory is an issue that we have already discussed in connection with the existence of psychological laws at §1.1 - if there are psychophysical laws, we might expect these states to be ultimately dispensable as neurophysiology developed.

\textsuperscript{33} Putnam makes this point in 'Other Minds'.

\textsuperscript{34} Versions of this claim are defended by Wittgenstein and Putnam. The most explicit defence is given by Strawson in \textit{Individuals}, University Paperbacks, Methuen, London, 1959. Wittgenstein held that such a claim was \textit{certain} in the sense that we could not rationally doubt it. He evidently thought that this precluded it from being a possible item of knowledge, since only that which could so be doubted could be known. I think that Wittgenstein is wrongly inflating a point about conversational procedures and implicatures into a substantive philosophical one. As Ayer remarked in commenting upon this view: "the fact that it may be pointless or even misleading for me to say such things as 'I know that looks red to me' or 'I know that these are my hands' in no way entails that what I am saying is not true" 'Wittgenstein on Certainty' p.231.
by such a pain-sufferer, etc. The Realist will claim that what guides the hearer’s grasp of the conclusiveness or otherwise of the evidence will be a conception of Jones being in a certain mental state - just that state the hearer is typically in when he encounters similar painful stimuli. Grasp of the classical truth-condition of ‘I am in pain’ governs the evidential verdicts the hearer makes - it is what determines his intuitions about the extent to which the evidence supports or fails to support the claim. Similarly, he will argue that the most perspicuous test for the hearer’s placing a classical interpretation on the truth of ‘I am in pain’ is the form of argumentation involving that statement that he accepts as valid. For a hearer could only acknowledge a classical argument as valid if he understood the conditions under which the assumptions, premises and conclusion were true.

The Anti-Realist is unlikely to be moved by any of this. Describing the hearer as having a grasp of a transcendent condition that governs his evidential verdicts as to the truth or falsity of ‘I am in pain’ is an over-rich description of the uncontested fact that the hearer knows the assertibility conditions of the sentence. Pointing to his endorsement of classical reasoning is illicit, he will contend, whilstever the use of such principles is sub judice since the Anti-Realist’s claim is precisely that those principles represent non-conservative extensions of their legitimate use in connection with decidable sentences.

It seems as if we have reached an impasse. But we should recall that there is in this particular case a strong presumption in favour of Realism, for several reasons. Firstly, Realism avoids the pitfalls of Semantic Variance - on the Realist’s showing, I can report the same event as Jones reports when he
utters 'I am in pain' by saying 'Jones is in pain'; secondly, that the postulation of an inner state responsible for Jones outburst and for his wincing and moaning explains the observable data whereas the competing 'anti-hypothesis' of the Anti-Realist - that these are simply primitive features of the language-game of sensation - simply refuses to allow that the data can be explained. Yet our whole practice of rationalising the behaviour of others and of ourselves depends upon such hypotheses. Moreover, the Anti-Realist's position seems dubiously coherent. Dummett's attempt to explicate Wittgenstein's 'non-referential' theory of the meaning of 'pain' seemed to presuppose the very fact the theory sought to deny - that 'pain' referred to the sensation of pain. Furthermore, as we saw in §2.4, the Anti-Realist's favoured replacement for the notion of a truth-condition - a 'criterion', as understood by Wittgenstein - appears to be, as both Wright and McDowell have persuasively argued it in fact is, incoherent. No sort of condition on 'I am in pain' or even 'Jones is in pain' can consistently combine the features of determining the sentence's meaning and permitting, on occasion, knowledge that Jones is in pain, if we also allow that the condition can, though satisfied at a time, later be defeated by the emergence of new evidence. Acceptance of criteria as semantic determinants of the meaning of 'I am in pain' seems to make knowledge of other minds literally impossible. For these reasons, I am inclined to think that the fact that we do understand the first person utterances of others provides a counterexample to the global Anti-Realist's claim that our understanding of any sort of statement is given by our grasp of its assertibility (or deniability) conditions.
§3.4 MANIFESTING UNDERSTANDING

The central issue to be examined in this section is whether a speaker can manifest a knowledge of classical truth-conditions and thus answer Dummett’s challenge.

I shall begin by looking at some unsuccessful attempts to answer this challenge. I shall then try to develop my own response.

§3.4.1 Realist replies to Dummett

In 'Truth and Use' and more recently in 'Realist semantics and content ascription', Colin McGinn has attempted to construct counter examples to the theory of content ascription implicit in Dummett’s writings. In his earlier article, McGinn imagined a community of intelligent immobile creatures living on the north side of a mountain for whom sentences about what was happening on the south side of the mountain were undecidable in principle. If certain recurrent features of their experience - such as the appearance and disappearance of sheep travelling from one side of the mountain to the other - could most reasonably be explained by postulating the existence of a south side to the mountain, what is to stop these creatures from so doing? Only an objectionable reductionist dogma of empiricism:
"the dogma... that our conceptual scheme cannot transcend our experience." McGinn replies. Moreover this Realist conception could be manifested by the tendency of these creatures to interpret the relevant assertions of fellow speakers: 'as expressions of the very realist beliefs we have seen no good reason to deny them.' This suffices to answer Dummett's challenge which should apply to this language if it applies at all, McGinn claims.

I am not convinced that this example is intelligible. All McGinn tells us is that his mountain-dwellers are like humans in their observational powers except that they are rooted to the spot. We are then supposed to just accept that they have a language with an undecidable fragment involving sentences about the south side of the mountain which they interpret realistically. Yet even if a Realist were to acquiesce in these stipulations without demanding some account of how the mountain dwellers' sentences were best to be interpreted in the absence of their possessing any clear analogue of agency and intention, it is certain that the Anti-Realist would not do so. For the fact is that these creatures' recognitional capacities are not the same as our own - sentences which we would find decidable are undecidable in principle for them. Hence, if meaning is determined by recognitional capacity, as the Anti-Realist asserts, it follows that their sentences do not and cannot mean

---

2 loc. cit. p.30.
the same as ours; we are therefore debarred from deciding whether their 'undecidable' sentences should be interpreted realistically or not\(^3\).

Apart from the issue of intelligibility, McGinn's diagnosis of and answer to the Anti-Realist's challenge are, I think, inadequate. It is question-begging to claim that their conceptual scheme can transcend their experience. For this is precisely the point at issue between the Realist and the Anti-Realist. McGinn's suggestion as to how such a Realist conception could be manifest in behaviour is also unacceptable as it stands - how does a mere propensity to interpret one another's assertions about the south side of the mountain as expressive of a realist belief about an inaccessible part of reality manifest an understanding of the relevant truth-conditions for those assertions? A Realist cannot just point to the prevalence of certain Realist-inspired practices in his attempt to meet the Manifestation Challenge. Since no one doubts that we often give expression to incoherent beliefs, the Realist has to show that Realism is not amongst these.

The Anti-Realist has two replies to this attempt to meet the Manifestation Challenge. The first reply is that the beliefs they form and impute to each other are incoherent. The mountain dwellers have, like us, wrongly assimilated the assertibility conditions and proof-theoretic relations of undecidable sentences to those of decidable sentences wherein assertibility

\(^3\)It is no use to appeal against this to the logical possibility of their sentences having the same meanings as ours since the Anti-Realist holds that the notion of logical possibility is grounded in meaning; indeed, his whole position rests upon the belief that logical laws such as Bivalence or Double Negation Elimination can be criticised on meaning-theoretic grounds. As the example does not present us with a possibility constructible in terms of our conceptual resources, it is incoherent. We should note, however, that the Anti-Realist is committed to conceptual relativism as his grounds for rejecting the intelligibility of McGinn's example reveal.
conditions and truth-conditions coincide. They have been misled as we have by the twin indefensible notions of truth as mere correspondence to Reality and meaning as truth-conditions.\footnote{Cf. Crispin Wright 'Truth Conditions and Criteria', loc. cit.}

The second reply is that even if, per impossibile, these beliefs were coherent, the mountain-dwellers would have no more chance of communicating them to each other than we have of communicating our beliefs about verification-transcendent states of affairs to one another - all either they or we actually communicate is the assertibility conditions of undecidable sentences.

The demand that meaning be exhaustively manifest in use is supposed to follow from the nature of meaning as an instrument for communication - this is supposed to require partitioning a speaker's semantic competence into its distinctive behavioural components.

At Elements of Intuitionism pp.377-378, Dummett offers the Realist a reply to the Anti-Realist which rejects the latter's demands for an exhaustive behavioural partitioning of semantic knowledge. This reply turns on the putatively theoretical character of meaning. I agree with Dummett that this is the most sophisticated Realist response he considers and I think that it is ultimately the right one to make, but the Realist must produce very compelling reasons for rejecting Dummett's Manifestation Constraint, since this Constraint is supposed to derive from the very nature of meaning as an instrument for communication. That meaning is essentially public and communicable is alleged to be the result of Wittgenstein's Private Language Argument. No reply to the Anti-Realist can in the end be sustained if this Wittgensteinian argument remains unanswered. It is the task of chapter 5 to examine the cogency of Wittgenstein's argument. The feasibility of the
response to Dummett that I develop in this section will thus be contingent upon the feasibility of my attempted rebuttal of the Private Language Argument in chapter 5.

Before attempting to develop the reply to Dummett which I favour, it might be useful to look at one other response to his argument which I believe also fails but for instructive reasons.

This response charges Dummett with too restrictive a conception of adequate manifestation of knowledge of truth-conditions. Very crudely, the claim is that Dummett restricts his acceptable models for a grasp of the truth-conditions of non-ED sentences to just two - the speaker's ability to explicitly formulate those truth-conditions in other terms and the ability to observe that a sentence's truth-condition is fulfilled when it is, but that if we look at a speaker's evidential and inferential practices, we will be able to ascertain whether he has hold of a concept of truth which transcends verification or falsification. With this type of response in mind, Dummett writes:

'It thus becomes conceivable that a certain model of meaning is required only in order to validate certain forms of inference, the employment of which is part of our standard practice. That is, that model of meaning would be unnecessary in order to account for the use of that fragment of the language which contained only sentences of a low degree of logical complexity ... And this would mean, therefore, that the meaning which, on such a model, we were taken as assigning to certain sentences, a meaning given in terms of their truth-conditions, was displayed only by our acceptance of certain forms of inference which could not otherwise be validated, rather than by anything involved in the use of those sentences as we learned it when, so to speak, they were on the frontier of the language we were acquiring (the frontier of complexity, that is).

It is just this which an opponent of a realist model of meaning finds incredible: he cannot believe that a grasp of a notion of truth transcending our capacities for its recognition can be acquired, and displayed, only by the acceptance of certain forms of reasoning. He concludes, instead, that these forms of reasoning, though generally accepted, are fallacious.'

Now the proponent of the approach Dummett is criticising here might respond by claiming that whilst our classically based inductive and deductive inferential behaviour does not itself explain whence we came

by the verification-transcendent conception of truth that governs the inferences which we make, it does nonetheless manifest such a conception, and since Dummett only requires a manifestation of such knowledge, he should concede that this sort of evidence fulfils his express requirements.

I think there is something right about this response, but I am certain that it cannot by itself answer Dummett. What I think is right is this: suppose we hold with Frege, as I urged in §1.2 we should, that truth is indefinable; then if speakers could be coherently attributed a grasp of the classical truth-conditions of non-ED statements, we should not expect any philosophically illuminating explanation of what having such a conception consists in - we should rather look to the evidence in their behaviour that they have this conception. On a Fregean understanding of truth, then, it makes sense to ask for evidence for our possessing a certain notion of truth even though neither in the behavioural evidence nor in anything else is there an explanation of what truth consists in.

But this position is quite consistent with the observation that the behavioural evidence is, as always, inconclusive and in need of radical interpretation. For while speakers might to a man all concur in their estimates of the degree to which certain evidence confirmed the (classical) truth of a statement or in their judgements as to which inferences are (classically) valid, this of itself is no guarantee that an explanation of these tendencies of theirs which made the best rational sense of their overall behaviour would not construe their grasp of the meanings of non-ED sentences as in reality a grasp of the assertibility conditions of those sentences (rather than a grasp of their classical truth-conditions). And, as we
know, this is precisely the Anti-Realist's claim - the attribution best supported by the behavioural evidence overall, the most parsimonious theoretically etc is one that credits speakers with a grasp of verifiable conditions rather than recognition-transcendent ones. The reason is simply that the ascription of a grasp of classical truth-conditions to a speaker is itself unverifiable, a speaker would never be able to definitively communicate his understanding of such conditions to any other speaker since he cannot in general recognise those conditions as obtaining when they do. Dummett nicely sums up the Anti-Realist response to the suggestion that we view classical inferential practices as by themselves manifesting a grasp of a Realist conception of truth in this passage:

'It is undoubtedly the case that if we have a grasp of some conception of truth ... with respect to which the principle of bivalence holds, then the laws of classical logic are valid; but it is hardly plausible that the mere propensity to reason in accord with those laws should constitute a grasp of such a notion of truth. If we consider any other class of statements for which it would be generally agreed that we do not possess a notion of truth subject to the principle of bivalence - for example, counterfactual conditionals - we can readily imagine that we had been induced, by childhood training, to apply the laws of classical logic to them, and we can recognise that, in such circumstances, we might be under a strong compulsion to suppose that we did have a notion of truth for such statements according to which each was determinately either true or false. ... Nevertheless, there seems no merit in the suggestion that, merely by undergoing training in applying the laws of classical logic to these statements, we should thereby acquire, what we now lack, a conception of truth for them under which each must be determinately either true or false.'

6 Elements of Intuitionism pp.376-377.
a grasp, that our inculcated inferential dispositions manifest a coherent conception of Realist truth rather than a socially transmitted illusion of such. So, we need a reason for thinking that we have such a coherent conception. I believe that we do have a reason for thinking this. To the extent that we grasp the meanings of first person psychological sentences as these are uttered by others, we have, or so I argued in the previous section, an understanding of the truth-conditions of those sentences. Moreover, the notion of truth applicable to those sentences seems to be, as Strawson claimed it in fact was, verification-transcendent. If I am right, the fact that we can understand the utterances of others as these pertain to their conscious experiences serves as a counter-example to Dummett's global Anti-Realism.

But is this right? The Anti-Realist will demand an exhaustive manifestation of our grasp of such sentences. I agreed in §2.4 that the general methodological demand for a behavioural manifestation of implicit knowledge is in general quite licit - where the Anti-Realist's interpretation of this requirement seems excessive is in its taking the requirement as tantamount to a demand for a behavioural reduction of each component of the practical ability, or, at the very least, for sentences about behaviour to form a reductive class for sentences attributing knowledge of truth-conditions of OL sentences to a speaker.

In the end, I will contend that this quasi-behaviourist demand can and ought in general to be rejected, but I stress again that the Realist is not at liberty to simply dismiss it just on the grounds that it is a quasi-behaviourist requirement - the Anti-Realist's claim is that it is forced upon us by the very nature of communication. For the moment, let us see how far we can go
toward meeting this constraint, which I will label (EMC) to designate the Exhaustive Manifestation Constraint. Our question then is 'Is there an exhaustive behavioural manifestation of a grasp of the meaning of the first person psychological utterances of others?'

Well, perhaps there is: a set of sentences might be a candidate for the class of first person psychological sentences in a language if the utterance of any sentence from that class by a speaker is treated as one whose truth-value the speaker is in a uniquely favourable position to decide. The behavioural signs of a grasp of the meaning of such a sentence amongst speakers will be that they ultimately do defer to the authority of the speaker in such instances.

It ought to be clear that this is not what the Anti-Realist is after: he wishes to correlate known truth-conditions for sentences with specific forms of behaviour which could plausibly be held to distinguish someone who understood those sentences from someone who did not. The behaviour mooted above for first person psychological sentences does not show of itself that speakers have hold of a coherent conception of truth for sentences from that class. To see this, we only need note that an Azande might argue on similar grounds for a unique behavioural manifestation of Azande speakers' grasp of the truth-conditions of an Azande witch-doctor's proclamations about the (undecidable) future - these are treated as authoritative where ordinary Azande conjectures about the future are not. This is the distinctive behavioural manifestation of an Azande's understanding of the truth-conditions of witch-doctor prophecies. Yet none of this shows that the competent Azande speaker has a coherent grasp of the truth-conditions for the witch-doctor's confident prophecy 'A village will never be built here'.
Now I argued in the previous section that the only plausible model for our grasp of first person psychological sentences is a truth-conditional one. But we have not vindicated that model as coherent and it seems that if the only way to do that is to correlate our understanding of such sentences with bits of behaviour which distinctively manifest that understanding, no such vindication looks possible at all.

Let us just pause and survey the dialectical situation.

(i) I argued that the Realist cannot hope to reply to Oummett by simply citing our classically-inspired evidential and inferential abilities as evidence for our possessing a grasp of Realist truth-conditions, for that behavioural evidence stands in need of radical interpretation and the Anti-Realist is free to reconstrue it as a confused grasp of the assertibility conditions of non-ED sentences.

(ii) I then pointed out that we had a class of sentences whose meanings seemed only to permit of a truth-conditional model in first person psychological sentences as uttered by others. This class of sentences could provide a counterexample to the Anti-Realist's claim that the only defensible model of the meanings of non-ED sentences is an assertibility conditional one if it could be shown that we had a coherent grasp of the meanings of those sentences.

(iii) But the problem was that the only way of showing this that was acceptable to the Anti-Realist was via a pairing of our grasp of the meanings of such sentences with distinctive behavioural manifestations, and there seemed no such distinctive behavioural candidates in the offing.

The task for the Realist is to provide a reason for thinking that the Anti-Realist's Exhaustive Manifestation Constraint (EMC) in (iii) above ought, if possible, to be rejected. Again, the appearance of 'if possible' is not idle, since as we recall (EMC) is supposed to follow from the very nature of language as a vehicle for communication.

Before attempting that task, let me interpose some remarks about what I think Dummett's arguments actually do establish. In my opinion, Dummett's arguments destroy a naive form of Realism defined by the twin doctrines that the meaning of a statement is given by its truth-conditions and that the truth of a statement is to be explained by its merely corresponding to
Reality. The reason that they destroy this form of Realism is that they destroy the latter doctrine - the truth of a statement is not from the Fregean, nor, I submit, from the Davidsonian, point of view explained by this notion of mere-correspondence-to-Reality, it is not explained by anything. If truth is taken to consist in such a correspondence, then it surely is appropriate to ask, as Dummett asks, how in general one recognises that such a correspondence holds if and when it does, or what in a speaker's linguistic behaviour could possibly display that his judgements as to when the correspondence held between a non-ED sentence and the section of Reality it purported to describe were generally reliable rather than wildly erratic. But if one sides with Frege and Davidson in jettisoning the Correspondence Theory as affording any explanation of truth, the need to provide an account of how we could bring ourselves into a position wherein we could recognise the truth-condition of a non-ED sentence as obtaining whenever it does will seem less pressing simply because it is not by virtue of its being paired off with a recognition-transcendent state of affairs that such a sentence is true if and when it is.

This, at any rate, is the crude metaphilosophical intuition. I do not intend it as a substitute for a hard argument against Dummett's Anti-Realist but only as an attempt to map out those versions of Realism which are knocked out by Dummett's argument and those versions (if, indeed there are any) which might yet survive his argument. On a holistic theory of meaning which does not construe the truth of sentences as consisting in any piecemeal pairing of sentences with discrete bits of reality, but which, taking truth as primitive, uses the whole truth theory for his language to model the speaker's whole practical linguistic ability, it is difficult to justify the demand for a
partitioning of this ability into the recognition routines Dummett requires in (EMC).

Dummett will reply that just because of this, holistic theories are intrinsically unacceptable, since (EMC) is forced on us by Wittgenstein’s Private Language Argument which he finds ‘incontrovertible’. Maybe so. But we might wonder whether the original demand for a pairing of known propositions of the meaning theory with separable practical abilities is not motivated instead by Dummett’s allegiance to his Principle C, what he takes to be the kernel of truth in the Correspondence Theory of Truth - namely that when a sentence is true, there must be something which makes it true, some knowable ground for its truth, a principle which, I have argued, Davidson is quite right to reject.

Dummett does provide a reason independent of Wittgenstein’s argument for believing that linguistic understanding is founded upon recognition abilities. It is this:

‘But to suppose that, in general, a knowledge of meaning consisted in verbalisable knowledge would involve an infinite regress: if a grasp of the meaning of an expression consisted, in general, in the ability to state its meaning, then it would be impossible for anyone to learn a language who was not already equipped with a fairly extensive language. Hence, that knowledge which, in general, constitutes the understanding of the language of mathematics must be implicit knowledge. Implicit knowledge cannot, however, meaningfully be ascribed to someone unless it is possible to say in what the manifestation of that knowledge consists: there must be an observable difference between the behaviour or capacities of someone who is said to have that knowledge and someone who is said to lack it.’

Dummett rehearses this argument in several different places and it is crucial to the success of his attack upon Realism. But what does it actually establish? Only, I submit, that semantic knowledge is in general implicit and that attributions of implicit knowledge must be constrained by the evidence of a

7‘The Philosophical Basis of Intuitionistic Logic’ p.217.
speaker's behaviour. But these are theses available to someone who denies on holistic grounds that those who know the meaning of an expression or sentence must evince a certain common pattern of behaviour.

The argument emphatically does not show, though I am not certain Dummett intends it to, that there is some class of words which can only be understood in a way that defies verbal explication and which has a uniquely non-linguistic manifestation in the behaviour of one competent in their use. Perhaps Dummett's argument secures the weaker thesis that for each speaker there must be some (not necessarily the same) class of expressions whose senses he cannot informatively state and for which the model of verbalisable knowledge is thus inappropriate? I submit that even this thesis is not warranted by the argument - a speaker might understand a word such as 'red' and be unable to informatively state its sense, yet it might be manifest to an interpreter from the speaker's verbal behaviour alone that he understands the term perfectly well insofar as he uses it in precisely the right verbal contexts.

From the fact that semantic knowledge cannot in general be assumed to be knowledge that a speaker can explicitly formulate, it does not follow that it is not verbalisable at all. An interpreter, acting as a 'Socratic midwife' could draw out in explicit form that implicit practical knowledge of the meaning of terms such as 'red' which is not readily verbalisable. How could he achieve this? By following Socrates' example and asking appropriate questions.

Now Dummett will object that this model is also of limited use as a model of understanding for it presupposes that the speaker understands the terms in which the questions are phrased, and how is the interpreter to test that supposition verbally?
My reply is that the supposition is tested not just by the verbal responses of the speaker but by his actions at large which is not to concede to Dummett that any one aspect of behaviour - the speaker's recognitional capacities or his words - should take explanatory precedence over any other. We test how (or whether) a speaker understands 'red' or 'square' by observing as best we can the 'conceptual role' the word occupies in the intentional organisation of the speaker's overall linguistic behaviour. Unless he understood 'red', certain utterances or responses to questions, certain discriminatory perceptual evaluations or executions of tasks, all patently presupposing that the agent must understand 'red', would appear totally inexplicable. But it might be manifest just from his verbal replies to our questions that he understands the predicate.

Dummett does suppose that his argument establishes the weaker thesis - that for each speaker, there must be some class of words which the speaker can only grasp in a non-verbalisable way. I have attempted to cast doubt on this supposition since I believe this model for implicit semantic knowledge to be deficient. Were truth to consist in a correspondence between sentence and an isolable state of affairs or fact, then, if the meaning of a statement is given by its truth-conditions, the fundamental model for linguistic understanding would be one of comparing the sentence with the relevant state of affairs in a recognitional act of some sort or other. But truth does not consist in such piecemeal correspondences and as a result this warrant for the recognitional model lapses.

Dummett has two methods of deriving (EMC). The first is through his argument above - that it follows from the nature of attributions of implicit
knowledge that such knowledge must be fully manifest in behaviour; the second through the Private Language Argument. I have argued that the first method fails. Whether the second succeeds depends upon whether the Private Language Argument goes through - chapter 5 is devoted to an examination of the most persuasive version of that argument of which I know: namely, Saul Kripke's.

Modulo this latter route to (EMC), there seem to be good reasons for rejecting this constraint. McDowell argues that (EMC) ought to be rejected because an interpreter only ever has finite stretches of verbal behaviour from which to determine what the speaker means by any given expression or sentence - thus, the finitely many applications of 'green' which R has been privileged to observe N make do not of themselves enable R to decide whether N means *green* or *grue* by 'green'.

Dummett can meet this initial criticism, I think. Consider the case of language-learning. McDowell's argument is that since all that N will ever have experienced in the language-learning situation is finitely many applications of 'green', R as his instructor cannot be certain that he has fastened onto the right property. But Dummett could reply to this that there are innate tendencies N shares with R which will in fact direct him to pick out the property of greenness rather than grueness. All that (EMC) asserts is that if N picks out the wrong property this will surface in his behaviour. But McDowell might reply that this is precisely what cannot occur - interpretations of even basic observational predicates such as 'green' which are, by our lights, deviant can only appear as such against a background of assumptions about the content of N's intentions, beliefs, desires and other

---

meanings which are of necessity underdetermined by the behavioural data. Dummett will have to assume that all or most of N's basic predicates are the same as R's, that his desires, intentions and beliefs are, at a certain level, similar to R's. Is this not too great a concession to the holist? Dummett might want to deny that it is - the only concession is that we all share a common nature, that our judgements of similarity are the same etc.; i.e. that some weak version of the Principle of Charity holds good for speakers of a language.

I cannot see how Dummett can make use of this assumption. What could it mean Anti-Realistically to assume that the experiences and attitudes of others match our own if not that we agree in our public use of sensation terms, colour words, in our verbal ascriptions of attitudes to others and to ourselves etc? This is surely the only content such an undecidable hypothesis could have for an Anti-Realist. But if this is the only content such a hypothesis can be given, it would be circular to invoke it in an effort to determine from his behaviour what a speaker meant by any given predicate.

In its most plausible form, (EMC) claims that a genuine disagreement between N and R over the meaning of a predicate or sentence must come to light in the divergent uses of the predicates or sentences that each makes. But I have been urging, along McDowell's lines, that this will only seem likely if we tacitly assume either detailed knowledge of all other (relevant) particular propositional attitudes N holds at a time together with the meanings of all other predicates relevant to the interpretation of the expression at hand or if we make a general assumption about N's probable psychological and semantic makeup. The difficulty for Dummett in either
case is that the (behavioural) evidence on which the given hypothesis is grounded can only be of use to us if it is interpreted - in which case we do not need the hypothesis - whilst the hypothesis can only be of use to us if the evidence is uninterpreted, in which case we can have no assurance that the hypothesis is even plausible let alone, as we need, true.

We can make the same point in a slightly different way. Dummett seeks, just as Davidson does, to discern from the gross behavioural evidence alone what N means by his words. But where Davidson insists that no interpretation of N's words can be achieved without a simultaneous interpretation of N's attitudes, Dummett thinks that at least the semantic determinant of N's linguistic behaviour will make a distinctive, repeatable and recognisable contribution to the causal production of that gross behaviour. It will do this because it will be recognisable whether N has grasped the assertibility conditions of the sentences of his language or not, this will be something R will, in principle, be able to determine. Yet even if it were fine-grained enough to settle what N means by his predicates and sentences, which it isn't, N's observed linguistic behaviour is indisputably a causal product of his attitudes and meanings, and to that extent offers at best equivocal support for any hypothesis that R might favour as to what certain of N's predicates or sentences mean. N's semantic agreements or disagreements with R can therefore never surface before R unalloyed by R's glosses on N's attitudes and other meanings. This by itself need not be a problem for Dummett. But the only plausible way to circumvent it - Davidson's way through the Principle of Charity - does create a difficulty for Dummett, since the Anti-Realistically acceptable
content to this principle casts it as a principle about speakers’ linguistic behaviour, the very thing we invoked the Principle to interpret!

Moreover, there is an additional twist here in the Realist/Anti-Realist dialectic if my argument in the previous section about the semantics of first person psychological statements was sound. I put forward that particular class of statements as a counterexample to a global assertibility conditional theory of meaning. The Anti-Realist could afford to cast doubt on the coherence of those statements given that the Realist could not produce a distinctive behavioural manifestation of a speaker’s understanding of the truth-conditions of those statements. But now it transpires that in order to even deploy (EMC) to rule out deviant interpretations of a speaker’s words, R must venture some sort of conjecture about N’s mental life, a conjecture for which those very statements would provide crucial evidence. Yet if the only plausible model for the meanings of those statements is a truth-conditional one, the Anti-Realist will be unable to avail himself of the conjecture at all.

So it seems that there are good reasons for either rejecting (EMC) or accepting semantic indeterminacy as the inescapable penalty of adhering to it. Meaning might not be able to transcend use, but an interpreter’s hypotheses as to what a speaker means most certainly will.

If we are justified in rejecting Dummett’s demand to correlate practical abilities on the part of a speaker in a piecemeal way with understood truth-conditions, there can no longer be, on holistic grounds at least, a serious question as to whether we do understand the truth-conditions of first person
psychological statements. Speakers are able to interpret, act upon and themselves produce such sentences, the assumption that others have minds of the same nature as our own being anyway rationally impossible to doubt. Our grasp of the truth-conditions of such utterances does not consist in our having a conception of what it would be like to be confronted by Jones' pain or sensuous experience, but in a quite general understanding of its being with Jones as it is with me which enables me to grasp the role such sentences play in the semantic economy of Jones' language. Such sentences, I have argued provide a counterexample to Dummett's global Anti-Realist semantics.

Pending an investigation of Wittgenstein's Private Language Argument, my provisional conclusion is that we can and ought to reject (EMC). But if we are not to pair meanings with distinctive behavioural manifestations in the way suggested by (EMC), in what sort of relation does a theory of meaning stand to the practical linguistic abilities of speakers' on the holistic model which I am recommending? Dummett himself nicely describes this model at Elements of Intuitionism pp.377-378:

'... the platonist is compelled to repudiate the principle that meaning is use ... he may choose to emphasise the *theoretical* character of a theory of meaning. Within such a theory, we explain a speaker's understanding of an expression or sentence by ascribing to him knowledge of some feature of it or by saying that he associates some semantic element or complex with it; but, on the present account, we do not then need to explain what it is for him to have this knowledge or make this association in terms of his linguistic behaviour. In constructing a theory of meaning, we are not, on such a view, attempting to articulate the complex of practical abilities that make up mastery of a language into its constituents, conceived of as isolable, though interconnected, practical abilities; we are merely aiming at what any theory attempts to provide, a picture which, taken as a whole, makes sense of a complex phenomenon, that is, makes it surveyable, even though there is no one-one correspondence between the details of the picture and observable features of the phenomenon. On this view, an acceptance of classical reasoning in mathematics does not constitute a grasp of a notion of truth for mathematical statements subject to the principle of bivalence ... rather, it warrants the ascription of a grasp of such a notion of truth to the individual concerned. This position does not represent a complete retreat into holism, since it allows the necessity of finding some theory of meaning, some general form of representation of that in which the understanding of a sentence consists, even though a theory of this kind does not need to be justified piece-meal.'
§4.1: INTUITIONISM

Platonism conceives of mathematical assertions as true or false insofar as they correspond or fail to correspond to a determinate, mind-independent reality. Intuitionism denies this thesis - the meaning of a mathematical sentence must be given in terms of those mental constructions that prove it, rather than in terms of (mysterious) platonistic states of affairs that happen to correspond to it. Mathematical sentences were to have their meanings explained by citing the grounds for asserting them, rather than states of affairs which made them true.

Brouwer adhered to a mentalistic (indeed a solipsistic) theory of meaning for mathematical statements. Dummett does not. It is highly likely, therefore, that the 'mental constructions' both appeal to in explaining the meanings of mathematical assertions are not exactly the same. Outside of characterising them metaphorically as mental constructions, the Intuitionist has no positive characterisation of the nature of natural numbers at all - unless he appeals to a Kantian intuition of the passage of time as Brouwer did.
Brouwer could conceivably have appealed to the apodictic nature of that intuition to have foreclosed the possibility that for some statement about the natural numbers, $s_k$, constructions $C_1$ and $C_2$ can be found proving both $s_k$ and $\neg s_k$. But Dummett clearly cannot do this. If such constructions $C_1$ and $C_2$ could be found, a primitive absurdity $\bot$ (such as '$0=1$') would become assertible, since $\neg s_k$ is assertible intuitionistically just when $\bot$ is derivable from the hypothesis of $s_k$. Hence there is a question as to how far traditional Intuitionism can be made consistent with the Anti-Realist's semantic strictures.

Proof-theoretically, intuitionistic logic is just a sub-system of classical logic, since every sequent valid in the former is valid in the latter, but not conversely. The principles of intuitionistic reasoning can be codified in an intuitionistic natural deduction system $I$ which has the following inference rules:

1 On Dummett's view, meaning must be exhaustively manifest in use, so no appeal to an apodictic intuition about the properties and structure of $\mathbb{N}$ that is not exhaustively manifest in the use one makes of statements expressing that intuition can be licit.
provided a does not occur in $\psi$ or in any of the assumptions on which $\psi(a/x)$ depends.

# provided a does not occur in $\psi$ or in $\delta$ or in any of the assumptions used in the derivation of $\delta$ from $\psi(a/x)$ except $\psi(a/x)$ itself.
We can define a translation-mapping $^*$ such that:

\[ 1^* = 1 \]
\[ \psi^* = \neg \psi \quad (\text{for } \psi \text{ atomic}) \]
\[ (\psi \land \theta)^* = (\psi^* \land \theta^*) \]
\[ (\psi \lor \theta)^* = \neg (\neg (\psi^* \lor \theta^*)) \]
\[ (\psi \rightarrow \theta)^* = (\psi^* \rightarrow \theta^*) \]
\[ (\forall x. \alpha)^* = (\forall x. \alpha^*) \]
\[ (\exists x. \alpha)^* = \neg (\exists x. \alpha^*) \]

Very roughly, the $^*$-mapping replaces wffs with their double negations. It can be proved that:

\[ \Delta \vdash_{\small \text{CPL}} \theta \iff \Delta^* \vdash_{\small \text{IPL}} \theta^* \quad \cdots (1) \]

where $\Delta, \Delta^*$ are sets of wffs and CPL, IPL denote classical and intuitionistic predicate logic respectively.

Similarly, from the classical semantic viewpoint, the topological structures used to interpret IPL (amongst which are the Beth trees and Kripke trees are special cases), appear as complicated classical structures with their own languages that can be interpreted in the ordinary classical model-theoretic way.

But from the Intuitionist's standpoint, these classical appraisals of intuitionistic logic and semantics are of dubious coherence. A completeness theorem for IPL proved in a classical metatheory will clearly not be faithful to the intended meanings of the intuitionistic logical constants. The intuitionist might claim that the $^*$-mapping shows CPL to be a weaker subsystem of IPL since a classical proof of a formula will be mapped onto an

intuitionistic proof of its double negation - e.g. a classical proof of $\exists x \alpha x$ will be mapped onto an intuitionistic proof of $\neg \neg \exists x \alpha x$.

But neither the classicist nor the intuitionist ought to be happy with an identification of classical wffs with their $\neg$-translations. For the fact is that they each mean different things by the logical constants - a classicist who asserts $\exists x \alpha x$ is not claiming that the supposition that $\exists x \alpha x$ cannot be proved leads to absurdity (which is how the intuitionist understands $\neg \neg \exists x \alpha x$), but that $\exists x \alpha x$ is true. The intuitionist for his part rejects such classical interpretations of the constants as incoherent.

As Dummett observes, the intuitionist's attitude to classically valid laws such as LEM and DNE will seem arbitrary until we see that it is based upon a complete rejection of Bivalence: $A \lor \neg A$ is only assertible when either $A$ can be proved or $\neg A$ can be proved - i.e. when $A$ is decidable.

Precisely what concept of truth (if any) the intuitionist can avail himself of is a very difficult question - Dummett searches for an interpretation of 'is true' more liberal than 'has been proved' (which has the consequence that any sentence not yet decided is neither true nor false), but not so liberal as to entail that there exists something which, were we to become aware of it,

---

3 Cf Dummett's remark at Elements of Intuitionism p.360: 'As Kreisel has emphasised, the intuitionistic philosophy of mathematics comprises two theses: a positive one and a negative one. The positive one is to the effect that the intuitionistic way of construing mathematical notions and logical operations is a coherent and legitimate one, that intuitionistic mathematics forms an intelligible body of theory. The negative thesis is to the effect that the classical way of construing mathematical notions and logical operations is incoherent and illegitimate, that classical mathematics, while containing, in distorted form, much of value, is, nevertheless, as it stands, unintelligible.'

4 'No account of the Intuitionistic rejection of LEM is adequate unless it is based on the Intuitionistic rejection of the Platonistic concept of mathematical truth as obtaining independently of our capacity to give a proof'. Elements of Intuitionism, loc. cit. p.18

5 I discuss the more general question of the concept of truth available to the Anti-Realist in S4.5.
we would regard as a proof (a notion Dummett finds unacceptably platonistic)⁶.

Apart from the intuitionistic biconditional (which can be defined as usual in terms of the conditional and conjunction), we can prove that no intuitionistic logical constant can be defined in terms of any of the others⁷. The divergence between the intuitionistic and classical understanding of the logical constants is most clearly in evidence over the quantifiers.

Classically, the only significant difference between domains of quantification lies in their cardinality. This is not the case intuitionistically - the intuitionist requires not just that a domain D of quantification be non-empty (¬∀x.¬x∈D), but that it be inhabited (∃x.x∈D) - i.e. that we can produce at least one element which we can positively assert belongs to D.

In addition, the intuitionist sharply distinguishes decidable from undecidable domains: a decidable domain being one specified by some finite description that allows us to recognise of any element whether it belongs to the domain or not. There can be no general requirement that a domain be decidable - a domain can be specified by any meaningful predicate whatsoever.

An example of an undecidable domain would be \(D_f = \{x \mid x=0 \lor (FLT \& x=1)\}\), where 'FLT' abbreviates Fermat's Last Theorem:— we could represent our possible states of knowledge with respect to \(D_f\) by a simple Kripke tree -

\[
\begin{array}{c}
\text{(0)} \\
\text{(0,1) - b - c} \\
\text{(0)}
\end{array}
\]

Since FLT has not been decided, we're in the state of information represented by node a: we know that 0∈D_f, but we do not know

⁶ Cf Elements of Intuitionism p.19.
⁷ For a proof of this result see Oxford Notes Toward the Formalisation of Logic, Parts III and IV, 1981, pp.180–181.
whether 1 belongs to \( D_f \) or not. Were FLT to be proved, we would be in the state represented by \( b \) - we'd know that \( D_f = \{0, 1\} \). Were FLT to be disproved, on the other hand, we would be in the state of information at \( c \) - we'd know that \( D_f = \{0\} \).

Dummett writes:
'the crucial assumption of classical logic is that the interpretation of the quantifiers remains the same, whether their domain be finite or infinite, denumerable or non-denumerable.'

The intuitionist's complaint with this assumption is that it erroneously assimilates infinite structures to finite ones - for, from the intuitionistic viewpoint, the infinite is precisely that which cannot be completed or surveyed in its entirety.

With finite processes, we can distinguish the outcome of the process from the process itself - distinct finite processes might generate a result which is recognisable independently of either process. But if we take an infinite structure such as the totality of natural numbers \( \mathbb{N} \), we can grasp that structure only insofar as we know its process of generation; a process which has no final result. Infinite structures are therefore of necessity incomplete and always in growth.

The Intuitionist's rather loose talk of 'constructions' is liable to mislead here. As Göran Sundholm has noted, intuitionists often equivocate between constructions-as-processes (\( \text{constructions}_p \)) and constructions-as-objects (\( \text{constructions}_o \)) in their use of the single term 'constructions'. Every mathematical object is a construction \( o \) and generated by a construction \( p \) but infinitary objects, unlike finite objects, can only be understood through the constructions \( p \) - the processes - that generate them. Thus, whilst each real

---

8 Elements of Intuitionism p.22.
number is extensional, it is incomplete, and can only be given to us through its corresponding generator which is intensional and complete\textsuperscript{10}.

According to Dummett, the Platonist regards the truth-value of a quantified sentence as if it were the outcome of a process of running through the values of its instances—on the assumption that the application of a well-defined predicate \( F \) to each element of the infinite domain has a determinate truth-value, true or false, the classicist concludes that its universal closure has a determinate truth-value, arrived at by forming the logical product of the values of its instances, and that its existential closure similarly has a determinate truth-value—the logical sum of the values of its instances. The intuitionistic objection to this thesis is that no mathematical or logical entity—truth-value or otherwise—can be the result of an infinite process for the simple reason that infinite processes have no finished result. Because an infinite domain can only be grasped intuitionistically as a process which generates the totality, when the domain of quantification is infinite, an existentially quantified sentence cannot be thought of as determinately either true or false, in advance of our possessing a proof of it\textsuperscript{11}, nor can a

\textsuperscript{10} Hence, the fact that reals are extensional does not entail that we can recognise them independently of their generators: by grasping the rule in accord with which the construction of a real number is determined, we might be able to recognise that two real number generators determine the same number. There remains a residual doubt as to whether the notion of an incomplete object is really coherent. But this doubt might stem from tacitly construing the term ‘object’ platonistically. If the intuitionist’s talk of ‘mental constructs’ is to be taken seriously at all, it would seem quite possible to have an incomplete ‘construct’.

In number theory, whilst we’re concerned with an infinite domain, \( \mathbb{N} \), the elements of the domain are finite objects; in analysis, however, the objects of the infinite domain \( \mathbb{R} \) are themselves infinite—real numbers generated by Dedekind cuts or Cauchy sequences or sequences of nested intervals of rationals. For the intuitionist, the real numbers are defined as species (the intuitionist’s intensional correlate to the classicist’s ‘set’) of equivalent real number generators. Insofar as their generating processes are infinite, the reals must correspondingly be infinite.

\textsuperscript{11} Or, more weakly, an effective method for proving it.
universally quantified sentence such as FLT be thought of as being true 'accidentally', as it were, i.e. each instance being true for a different reason. It is not clear to me that the classicist really does think of the infinite in the manner that the intuitionist alleges, since he sharply distinguishes the process whereby we come to know of the domain from the domain itself. In the case of spatial and temporal continua which the classicist alleges have the cardinality of the classical continuum, there is nothing to be found in human experience - no 'smallest perceivable spatial (or temporal) point' - with which the points of these continua could be paired. Hence it cannot be right to assimilate the classical conception of such non-denumerable physical domains as one wherein an interminable process\textsuperscript{12} of determining the values of the instances of a sentence quantified over that domain somehow, mysteriously, comes to an end. For the classical spatio-temporal continua are precisely not 'in growth' - rather, they are conceived as literally consisting of a non-denumerable number of unextended spatio-temporal points.

The intuitionist will no doubt respond that our only grip upon the notion of the infinite arises through our associating it with an interminable process. So, the classicist can have no coherent conception of the continuum at all. But there is an important theoretical consideration that lends powerful support to the coherence of the classical continuum. This has to do with Zeno's Paradox of Plurality and its resolution.

Zeno argued that if extended things exist, they must be composed of parts, which in turn have parts. Because the process of subdivision is indefinitely repeatable, there must therefore be an infinity of parts. Now either the

\textsuperscript{12} The process would be interminable because there are infinitely many objects, not because each instance requires an interminable period of time to determine its truth-value.
ultimate parts have magnitude or they do not. If they do have magnitude they can be further subdivided, so they must have no magnitude. But no extended object can be composed of parts of zero magnitude. So the parts must have magnitude. But the addition of an infinite number of magnitudes all exceeding zero will yield an infinite magnitude. Hence, extended objects, if they exist must both be without magnitude and infinitely large.

Adolf Grünbaum has resolved this paradox using the resources of a measure theory which assumes the existence of a non-denumerable continuum. As Grünbaum points out, anyone who rejects non-denumerable totalities as constructivists do, must develop an alternative measure theory in which the measure of the continuum is not zero. It is not just that the classical continuum is vindicated as coherent on Grünbaum’s theory, it is alleged that physical spatio-temporal continua have the cardinality of that continuum. Modern physics is predicated on a classical interpretation of the quantifiers. For the intuitionist, the paradox only arises because of the mistaken assumption that the continuum is composed of ultimate parts – i.e. points. The intuitionistic continuum is always in growth, real numbers being generated by free choice constructions as well as rule-governed operations (e.g. solutions to linear equations). The classical question as to what cardinality this continuum has simply fails to make sense on such a model.

Yet the intuitionist’s charge was that the classical conception of the infinite is incoherent, but Grünbaum, in resolving Zeno’s paradox of plurality,

14 Hermann Weyl’s conception of the continuum as developed in *Das Kontinuum* in which the only real number generators are those corresponding to algebraic numbers leads to a continuum of denumerable cardinality and would thus seem open to Zeno’s Paradox.
ostensibly provided this continuum with foundations that are logically coherent from the classical point of view.

Should the matter simply rest there with a dialectical stalemate? I do not think it should for if modern physics really is, as it certainly seems to be, predicated on a classical conception of the continuum, and if, as Grunbaum argues, spatial and temporal continua are classical continua, then how could it arise that a conceptually incoherent model should be so astonishingly successful in practice? Moreover, the intuitionist owes us an explanation of the cardinality of physical continua, for even if he can evade Zeno's Paradox for the mathematical continuum by maintaining it is always in growth, it is quite uncertain what account he should give of the cardinality of a stretch of space which is not 'in growth', at least not in the same way. Pending no satisfactory answer here, the Paradox has only been forestalled. Surely, the classicist could insist, Quine is right to hold that the success of total theory gives us the best possible reason to believe in its theoretical posits and surely success is only possible where the theory is coherent?15

How then does the Intuitionist understand \( \forall x.F(x) \) and \( \exists ! x.F(x) \) where \( x \) ranges over an infinite domain? For simplicity's sake, take the domain of

---

15 Unless of course that practice is shot through with Realist assumptions that are ultimately untenable on apriori meaning-theoretic grounds alone, as Dummett claims. I think the epistemological and metaphysical considerations by themselves strongly favour the classicist, if Grünbaum is right about the cardinality of physical continua. So whilst I am, in one way, simply endorsing Dummett's argument in 'The Philosophical Basis of Intuitionistic Logic' that a cogent case for relinquishing classical logic and mathematics has to be based upon general meaning-theoretic grounds, I think, contrary to Dummett, that prior to reviewing such Anti-Realist considerations the situation actually favours the Realist, which is not to say that the epistemology of mathematical objects may not, for Benacerraf-styled reasons, pose an extremely difficult problem. Quine, I think, would approve of Dummett's verificationist tendencies but deplore the jettisoning of classical logic, again adducing holistic methodological reasons for his conservatism here - i.e he would claim that one cannot make any sensible decision as to which logic or 'meaning-theory' to choose on apriori grounds alone. Whether this is defensible naturally depends on the tenability of his holism. At the very least, I think it should be granted that if Grünbaum is correct, the intuitionist's complaints about the coherence of the classical conception of the continuum are toothless outside of the context of Dummett's meaning-theoretic arguments against Realism.
quantification to be \( \mathbb{N} \) and let \( F \) be a decidable predicate. Our domain is
generated by the operation of forming the successor of each number starting
at 0. Since \( F \) is a decidable property of the natural numbers, there exists a
second process which generates truth-values in the following manner: \( F_a \) is
true when the object denoted by \( a \) that is formed at the corresponding stage
in the process that forms the domain, has \( F \); \( F_a \) is false otherwise. Then
\( \forall x. Fx \) is to be interpreted as asserting that this second process coincides
with a third process generating only truths \( < T, T, T, \ldots > \).
For \( \exists x. Fx \), we consider a third process generating only falsehoods \( < F, F, F, \ldots > \)
and \( \exists x. Fx \) asserts that our second process does not coincide with this one.
So, \( \forall x. Fx \) means that the rule governing the second process guarantees that
it yields the sequence \( < T, T, T, \ldots > \), and \( \exists x. Fx \) that the rule yields some
sequence other than \( < F, F, F, \ldots > \).

Let us see how the differing interpretations of the constants manifest
themselves in the mathematical practice of the classicist and of the
intuitionist. It might be thought that the intuitionistic constants could be
embedded in a classical fragment:-
Thus we could classically define a constructive operator \( 'U' \), such that \( 'AUB' \)
meant:
'we have a constructive proof of \( "Av B" \) ... (i)
Similarly, we could define a classical constructive operator \( '3' \), such that
'\( 3x. Ax \)' meant:
'we have a constructive proof of \( "\exists x. Ax" \) ... (ii).

Although the intuitionist will reject (i) and (ii) as unintelligible, this
complaint would have no force anti-realistically if the observable practice of
the classical mathematician interested in constructive proof differed not at
all from his intuitionistic counterpart.
But Dummett contends that we can see that this is not so from the intuitionistic invalidity of (the classically sound) Markov’s Principle\(^\text{16}\):

$$\forall x (A(x) \lor \neg A(x)) \land \neg \forall x \neg A(x) \rightarrow \exists x A(x) \quad \ldots \text{(MP)}.$$  

\(\text{(MP)}\) is still assertible classically even when we replace \(\exists\) by \(\exists\) and \(\lor\) by \(\lor\). For, Dummett argues, \(\forall x (A(x) \lor A(x)) \land \neg \forall x \neg A(x) \rightarrow \exists x A(x) \) \(\text{(MP')}\) expresses the classical truth that:

\(\ldots \text{if we have an EM for deciding of every } n \text{ whether } A(n) \text{ or } \neg A(n) \text{ and we know that not every number falsifies } A(x), \text{ then we can actually find a number that satisfies it.} \)\(^{17}\)

since if \(\neg \forall x \neg Fx\) holds, the process of running through each number in the domain 0, 1, 2,\ldots and testing whether it satisfies \(Fx\), breaking off as soon as we find one that does, must terminate.

But the corresponding intuitionistic interpretation of \(\text{(MP)}\) generates an intuitionistic falsehood: for \(\neg \forall x \neg Fx\) does not mean, intuitionistically, \(Fx\) will not as a matter of fact be found to fail for every natural number, as it does classically, but rather: ‘the supposition that \(Fx\) fails for every natural number leads to absurdity’, which does not guarantee that by testing each natural number in turn we shall eventually chance upon one satisfying \(Fx\).

So, Dummett contends, there is no special classical constructive reading of \(\exists\) and \(\lor\), alongside a more general non-constructive reading which the intuitionistic reading of these constants captures\(^{18}\).

The intuitionist claims that both constructive and non-constructive classical readings are senseless: \(\exists x Fx\) \textit{can only mean that} we have a proof (or can provide an EM for proving) of a specific element of the domain that it

\(^{16}\) loc. cit. p. 21  
\(^{17}\) loc. cit. pp. 21–22.  
\(^{18}\) loc. cit.
satisfies $Fx$. Hence for the intuitionist, the truth-tables do not render the classical meanings of the constants intelligible and the crucial assumption required to effect this - vis, that every statement is determinately either true or false - must be rejected.

Assuming that we understand what it is for a construction $C$ to be a proof of a mathematical wff $\Theta$\textsuperscript{19}, we can give a recursive definition of the proof-conditions of a complex wff in terms of those of its parts following Heyting's informal exposition\textsuperscript{20}. For simplicity's sake, we again assume that the variables of our language range over the natural numbers:

(i) Construction $C$ is a proof of $(\Theta \& \Psi)$ iff $C$ is a pair $(C_1, C_2)$ such that:

$C_1$ is a proof of $\Theta$ and $C_2$ is a proof of $\Psi$.

(ii) Construction $C$ is a proof of $(\Theta \vee \Psi)$ iff $C$ is a pair $(C_1, C_2)$ such that:

$C_1=0$ and $C_2$ is a proof of $\Theta$ or $C_1=1$ and $C_2$ is a proof of $\Psi$.

(iii) Construction $C$ is a proof of $(\Theta \rightarrow \Psi)$ iff $C$ is a construction that converts each proof $p$ of $\Theta$ into a proof $C(p)$ of $\Psi$.

(iv) No construction is a proof of $\bot$.

(v) Construction $C$ is a proof of $\forall x. \Psi x$ iff $C$ is a pair $(c_1, c_2)$ such that:

$c_1$ is a proof of $\Psi_{c_2}$.

(vi) Construction $C$ is a proof of $\forall x. \Psi x$ iff $C$ is a construction such that:

for each natural number $n$, $Cn$ is a proof of $\Psi_n$.

\textsuperscript{19} The clause for the atomic case will depend upon the theory under discussion - for the case at hand, there is no problem because the closed atomic wffs are all decidable arithmetical equalities.

\textsuperscript{20} The version I give here is similar to van Dalen's op. cit. p.231.
Kreisel and Goodman have incorporated these intuitive explanations of the constants into a mathematical theory of constructions. Dummett believes that whilst their theory is: 'as yet still in an imperfect state of development... there is no doubt that these standard intuitive explanations of the logical constants determine the intended intuitionistic meanings, so that anything which can be accepted as the correct semantics for intuitionistic logic must be shown either to incorporate them, or, at least to yield them under suitable supplementary assumptions.'

There is a prima facie ground for doubting the adequacy of these intuitive explanations which Dummett acknowledges: they are highly impredicative inasmuch as they require us to ascertain what effect a construction will have when applied to an arbitrary proof of a conditional or to an arbitrary proof of a negated sentence before we can determine whether the construction genuinely is a proof of the relevant conditional or negated sentence. As this seems to require that we must in some sense be able to grasp a totality of constructions for a given conditional or negated sentence that includes all possible proofs of such sentences, these explanations might appear to be viciously circular.

Another feature of the Heyting explanations which is more easily resolved is their apparent lack of fit with the actual practice of intuitionists. Disjunctive statements often are asserted in their writings without a proof of one or the other disjunct - it is common practice to assert \( A \lor B \) when one merely has an EM for obtaining a proof of either \( A \) or of \( B \) (e.g. when \( A \) is a decidable statement and \( B = \neg A \)).

An informal proof of \( A \lor B \) (or of \( \exists xFx \)) which merely cites an EM for finding a proof of either disjunct (or for finding an object \( k \) such that \( Fk \)) should not in the strict sense be regarded as a proof, but only as an EM for finding one.

---

21 Elements of Intuitionism p.389.
22 loc. cit. p.390.
Dummett contends that we should acknowledge a distinction between strict or canonical proofs and informal demonstrations which supply EMs for constructing canonical proofs. The first step in making Heyting's informal semantics a rigorous explication of the meanings of the intuitionistic logical connectives will then be to replace each occurrence of 'proof' throughout with 'canonical proof', according to Dummett.

The intuitionist has a means of making the meanings of his connectives clear to the classicist in the following way:

- $\boxdot(\Theta \& \Psi)$ asserts that $\boxdot\Theta \& \boxdot\Psi$
- $\boxdot(\Theta \lor \Psi)$ asserts that $\boxdot\Theta \lor \boxdot\Psi$
- $\boxdot(\Theta \rightarrow \Psi)$ asserts that $\boxdot(\boxdot\Theta \rightarrow \boxdot\Psi)$
- $\boxdot(\neg\Theta)$ asserts that $\boxdot(\neg\boxdot\Theta)$

The connectives to the left of 'asserts that' are the intuitionistic ones; the connectives to the right, their classical counterparts. $\boxdot$ means 'we have a proof that'.

---

23 As we shall see, the success of an anti-realist theory of meaning depends in no small part on the viability and the generality of the distinction between canonical proofs and demonstrations. For Dummett hopes to generalise the distinction between canonical proofs and demonstrations so that it applies to all types of sentences of a language in the form of a distinction between a direct means of verifying a statement and an indirect means where the possibility of a direct means of verifying a statement is grasped by any speaker who understands the statement. The problem for Dummett is that his most compelling argument for the canonical proof/demonstration distinction is, or so I shall argue, unsound.

24 Dummett argues that the alternative proposal - that we so modify the intuitive explanations that we permit C to be a proof of $A \lor B$, for example, if it provides an EM for finding a proof of either disjunct - seems less attractive once we examine the informal explications of the meanings of $\lor$, $\rightarrow$, $\neg$. 

We can now provide a translation of any ISL wff into a wff of S4 by working from the outside inwards and eliminating every intuitionistic constant in favour of its classical counterpart. Call $\Theta^*$ the modal translation of an ISL wff $\Theta$. Then:

- $P^* = \Box P$ (P atomic)
- $(\Theta \& \Psi)^* = \Theta^* \& \Psi^*$
- $(\Theta \lor \Psi)^* = \Theta^* \lor \Psi^*$
- $(\Theta \rightarrow \Psi)^* = \Box (\Theta^* \rightarrow \Psi^*)$
- $(\neg \Theta)^* = \Box (\neg \Theta^*)$

We can then prove that:

A formula of ISL is intuitionistically valid iff its modal translation is S4-valid.

Moreover, we can give an argument for the failure of LEM similar to the one a supporter of S4 might give to reject the definitive S5 axiom: $\Box A \Box \neg A$, for the modal translation of $A \lor \neg A$ is just that axiom. Reading $\Box A$ as 'we have a proof of $A$', the S5 axiom states that: 'it is either the case that we have a proof of $A$ or that we can prove that no proof of $A$ will ever be forthcoming', which is clearly false for currently unsolved mathematical problems such as Fermat's Last Theorem.
§4.2. PROOF THEORY AND ANTI-REALISM.

Dummett's main argument for revising our classical inferential practices depends crucially on certain proof-theoretic ideas and results which Prawitz and Dummett have developed between them. In this section, I will be chiefly concerned to bring out those of Prawitz's results which bear upon this argument.

Recall from our discussion of holism in §2.5 that Dummett endorses Quine's use of the image of a web of interconnected sentences as a theoretical model for our understanding of our language. As we shall see in §4.5, this model is in fact essential for Dummett's own account of verification\(^1\). The understanding of a single sentence will in general presuppose an understanding of a fragment of the language which may be quite extensive. Unlike Quine's holistic model, however, no sentence will feature in that fragment which does not contain components of the given sentence. Dummett therefore denies the essential doctrine of semantic holism - that any given sentence has inferential relations with every other sentence in the web; on the contrary, we can impose a definite structure on the web that will approximate at least a quasi-ordering. From this perspective, holistic theories of meaning are indeed structureless.

We can now offer a rough outline of Dummett's argument against classical reasoning:

\(^1\) Although both start from the same model, they develop it in very different ways.
(1) Our understanding of our language may be represented as (at least) a quasi-ordered structure, the order being determined by our progressive acquisition of the sentences of the language.

(2) What we actually acquire in acquiring our language is a knowledge of the assertibility conditions for its sentences. Insofar as we have a conception of truth for these sentences, it is constrained by what we recognise in practice as establishing truth.

(3) Thus, the concept of truth applicable to the atomic fragment of the ordered structure will be a recognisable one - something like intuitionistic truth.

(4) Hence, the concept in terms of which the deducibility relation between sentences is to be characterised will be (something like) intuitionistic truth.

(5) But, relative to such a concept, the use of classical modes of reasoning in relation to the atomic fragment induces a non-conservative extension of that fragment.

(6) Upon that basis, we can criticise those classical practices and seek to revise them - for the conditions for asserting a sentence can no longer be expected to be in harmony with the consequences that can be drawn from its assertion once we are able to infer sentences by those classical rules that we were previously unable to infer prior to their introduction.

Clearly, (2), (3) and (4) constitute the crucial premises of the argument. So how does Dummett hope to establish that the concept of truth speakers actually acquire through their training in the use of the sentences of a natural language is a non-classical one? Dummett tells us at William James Lectures (WJL) 3:7 that disputes about the general validity of logical laws are to be settled either by seeing which laws are justified by the meaning theory or by determining which of two rival theories of meaning is correct. Clearly, it is only the second of these two methods which is applicable here. So, (2), (3) and (4) will depend upon the success of Dummett's meaning-theoretic argument against Realism.

2 Cf 'Philosophical Basis of Intuitionistic Logic' p.222: 'It will always be legitimate to demand, of any expression or form of sentence belonging to the language, that its addition to the language should yield a conservative extension; but, in order to make the notion of a conservative extension precise, we need to appeal to some concept such as that of truth or that of being assertible or capable in principle of being established, or the like; and just which concept is to be selected, and how it is to be explained, will depend upon the theory of meaning that is adopted' [italics mine]
Yet, if this is Dummett's strategy for establishing (1), there is an obvious flaw in it: for one natural Realist reply to Dummett's demand for a practical manifestation of a speaker's knowledge of the truth-conditions of non-ED sentences was to cite our classical deductive practices involving those sentences as evidence that we'd grasped a classical conception of truth for them. I agreed that our practices did not constitute a grasp of classical truth but insisted that it might yet manifest such a grasp.

It would therefore be patently question-begging on Dummett's part to presume the success or his Manifestation Challenge in order to mount an argument for the revision of a practice which the Realist contends meets that challenge. Dummett must therefore avoid the threat of vicious circularity in his strategic argument for the revision of classical modes of reasoning.

As we noted in the previous section, the introduction and elimination rules in a natural deduction system I run parallel with Heyting's explanations of the meanings of the logical constants. Gentzen hinted at an idea which Prawitz and Dummett have sought to make precise when he wrote:

'an introduction rule gives, as it were, a definition of the constant in question... an elimination rule is only a consequence of the corresponding introduction rule, which may be expressed somewhat as follows: when making an inference by an elimination rule, we are allowed to "use" only what the principal operator of the major premises "means" according to the introduction rule for this sign.'

Can the introduction rules in a ND system really be thought of as defining the meanings of their associated constants? One logician unconvinced that they could was Arthur Prior. Prior invented a propositional connective 'TONK' with the following introduction and elimination rules:

---

3 Peter Menzies brings this out clearly at ch.3 of his Realism: A Critical Examination of some arguments of Michael Dummett's, M.Phil Thesis (unpublished), University of St. Andrew's, 1980.

Prior showed that it was possible, using TONK I then TONK E, to prove false conclusions from true premises - for example, we could prove the absurd thesis that any two arbitrary propositions A and B are identical:

\[
\begin{array}{c}
\text{(TONK I)} \\
\hline
A \quad \text{TONK} \quad B \\
\end{array}
\]

\[
\begin{array}{c}
\text{(TONK E)} \\
\hline
A \quad \text{TONK} \quad B \\
\end{array}
\]

There must therefore be some constraints placed upon an introduction rule before it can serve as definiens for some logical constant.

In response to Prior's TONK, Nuel Belnap urged that there was a purely proof-theoretic way of ruling out such operators as illicit - adopting TONK's inference rules could be shown to be inconsistent with accepting certain antecedent assumptions about deducibility. Belnap expressed this point by saying that TONK induced a 'non-conservative extension' of the original characterisation of deducibility.\(^5\)

Now while the term 'conservative extension' is frequently applied to certain types of relations between languages or theories, as Dummett and Prawitz use this notion and as Belnap originally intended it, a conservative extension is a relation between logics. We can formulate the requisite condition thus:

---

Let $\lambda_0$ be a logic formulated in language $L_0$ and consisting of inference rules $R_1, R_2, \ldots, R_k$ and axioms $A_1, A_2, \ldots, A_m$ (this could be a null set as in a ND system), then if, upon the addition of inference rules $R_{k+1}, \ldots, R_n$ and, if any, axioms $A_{m+1}, \ldots, A_p$ to $L_0$, there is no set of sentences $\Gamma$ of $L_0$ and sentence $p$ belonging to $L_0$ such that $p$ is derivable from $\Gamma$ in the extended logic $\lambda_1$ but not derivable from $\Gamma$ in $\lambda_0$, then $\lambda_1$ is said to be a conservative extension of the logic $\lambda_0$.

More succinctly:
\( \lambda_1 \) will be said to be a **conservative extension** of \( \lambda_0 \) just when any sequent \( q \) that holds in the expanded system \( \lambda_1 \), if both premises and conclusion belong to the original language \( L_0 \), is already a sequent of \( \lambda_0 \). 6

An elimination rule can be seen as the inverse of the corresponding introduction rule in the sense that a proof of the conclusion of an elimination is already available if the major premise of the elimination is inferred by the corresponding introduction rule. 7 This principle, due to Prawitz, is known as

---

6 Belnap’s account starts from the assumption (which is implicit in both Prawitz’s and Dummett’s use of the notion) that we have some characterisation of deducibility that completely determines the context of deducibility, one that we can treat as a formal system in its own right – i.e. as a set of axioms and rules involving the deducibility sign ‘\( \vdash \)’. Belnap chose ‘for definiteness’ Gentzen’s structural rules of Weakening, Contraction, Permutation and Transitivity. The formal system is supposed to express ‘all and only the universally valid statements and rules expressible in the given notation: it completely determines the context’ [‘Tonk, Plonk and Plink’ p.46 in Copi, I and Gould, J (eds.) loc. cit.].

A proposed definition of some connective PLONK will then, according to Belnap, be an extension of the formal system in two senses:

(a) the notion of sentence is extended by introducing \( A \cdot \text{plonk} \cdot B \) as a sentence, whenever \( A \) and \( B \) are sentences. (b) We add some axioms or rules governing \( A \cdot \text{plonk} \cdot B \) as occurring as one of the premises or as conclusion of a deducibility statement. These axioms or rules constitute our definition of plonk in terms of the role it plays in inference. ... We may now state the demand for the consistency of the definition of the new connective, plonk, as follows: the extension must be conservative, i.e., although the extension may well have new deducibility-statements, these new statements will all involve plonk. The extension will not have any new deducibility-statements which do not involve plonk itself. It will not lead to any deducibility-statement \( A_1, A_2, \ldots, A_n \vdash \)

\( B \) not containing \( \text{plonk} \), unless that statement is already provable in the absence of the \( \text{plonk} \)-axioms and \( \text{plonk} \)-rules. loc. cit. pp.46-47.

It is clear that this is substantially the same explanation of the notion of a conservative extension as given above, for Belnap’s ‘characterisation of deducibility’ is simply an attempt to characterise a logic. I stress that the relation of conservative extension is a relation that holds between *logics* because there are different senses of the term ‘conservative extension’ – e.g. Schoenfield J., *Mathematical Logic* 1941, describes it as a relation between theories: ‘A conservative extension of (a theory) \( T \) is an extension \( T' \) of \( T \) such that every wff of \( T \) which is a theorem of \( T' \) is also a theorem of \( T \).’

The important point which Belnap insists upon is that the characterisation of deducibility implicitly assumed must be taken to be *complete* prior to the addition of the given inference rules to the language \( L \), otherwise there will be nothing objectionable in the fact that we can deduce statements by means of the new rules which we could not deduce before.

the Inversion Principle and can be formulated more precisely in the following way:

Let \( \mathcal{L} \) be a logical constant and \((\mathcal{L}I)\) and \((\mathcal{L}E)\), respectively, be the introduction and elimination rules for \( \mathcal{L} \). Then if \( C \) is a consequence of applying the elimination rule \((\mathcal{L}E)\), then a deduction of the major premise of \((\mathcal{L}E)\) whose last step consists of an application of \((\mathcal{L}I)\), already contains a deduction of \( C \).

As an illustration, consider Modus Ponens \((\rightarrow B)\): The introduction rule for \( \rightarrow \) stipulates the condition for inferring \( B \rightarrow C \):

\[(\rightarrow I) \quad B \quad \vdash B \rightarrow C\]

\( B \rightarrow C \) can be inferred whenever we have a proof of \( C \) available from the hypothesis of \( B \).

\[\vdash B \rightarrow C \quad \vdash C\]

Modus Ponens, the elimination rule \((\rightarrow E)\) \( B \quad \vdash B \rightarrow C \quad \vdash C \)

is the inverse of \((\rightarrow I)\) in the sense that if the condition for inferring \( B \rightarrow C \) as stated by \((\rightarrow I)\) is satisfied (i.e. there is a proof of \( C \) from the hypothesis of \( B \)), then a proof of \( C \) is already available by replacing every hypothesis \( B \) by the given proof of \( B \). The deduction of the major premise \( B \rightarrow C \) thus already contains a deduction of \( C \), the consequence of the elimination inference.

Consecutive applications of \((\rightarrow I)\) and \((\rightarrow E)\) allow us to deduce from a set of sentences \( \Gamma \) only those sentences which were deducible from \( \Gamma \) without those rules.

It is clear that Prior's TONK induces a non-conservative extension if it is added to a set of sentences since successive application of \((\text{TONK I})\) and \((\text{TONK E})\) permit the derivation of \( A \rightarrow B \), a sentence which is not deducible without the aid of those inference rules.

---

The Inversion Principle gives rise to certain reductions which apply whenever a sentence obtained as the conclusion of an introduction rule is then used as the premise of an elimination rule. The reduction removes this 'loop' in the proof. In our example:

\[ \begin{array}{c}
B \\
\vdots \\
C \quad \text{(→I)} \\
\vdots \\
B \Rightarrow C \\
\vdots \\
C \quad \text{(→E)} \\
\end{array} \]

That is, we can deduce C without the detour of first introducing \( B \Rightarrow C \), then applying (→E). When all such reductions have been carried out, the proof is said to be in *Normal Form* - a normal derivation proceeds directly from certain hypotheses to its conclusion without introducing any sentences in the proof that are not deployed to build up the conclusion of the proof\(^9\).

A Normal Form Theorem states that if \( \Delta \vdash \varphi \), then there exists a normal derivation of \( \varphi \) from \( \Delta \). A Normalization Theorem states that any derivation reduces to a normal derivation where the relevant notion of reduction is as above. A Strong Normalization Theorem states that every sequence of reduction steps terminates in one in normal form.

Dummett requires that for all sentences of a natural language there be a harmony between two identifiable aspects of their use:

1. **The conditions** under which their assertion is warranted.

---

\(^9\)An important corollary of the normal form of a derivation is the Subformula Property: in a normal derivation of \( \Delta \vdash \varphi \), only subformulas of \( \Delta \) and \( \varphi \) can occur in the derivation. This result is important because it then follows directly that Intuitionistic Predicate Logic (IPL) is a conservative extension of Intuitionistic Sentential Logic (ISL). For if we let \( \varphi \) be a proposition which is a theorem of ISL (\( \vdash \neg \varphi \)), and consider a normal derivation \( D_\varphi \) of \( \varphi \) in IPL, we know from the subformula property that only propositional connectives can occur in this derivation. Hence, we have a derivation of \( \varphi \) using only propositional rules, and thus IPL is a conservative extension of ISL.
(2) The consequences which flow from their assertion.

This broader requirement generalises the proof-theoretic constraint which holds between the introduction and elimination rules for first order logic\textsuperscript{10}. There is good reason to believe that the constraint can be imposed upon constructive mathematical sentences: Per Martin-Löf has analysed intuitionistic mathematical proofs with inductive and iterated inductive definitions into introduction and elimination inferences and proved a corresponding normalisation theorem\textsuperscript{11}. Martin-Löf has formalised a considerable fragment of constructive mathematics by means of an intuitionistic theory of types which he has developed for the task. In his system, the notion of introductions and eliminations applies not just to logical constants but to atomic predicates as well.

But outside of a mathematical or logical context wherein the notion of proof can be clearly defined, can we legitimately demand that a proof-theoretic constraint be met? Does the constraint even make sense in the wider context? Although sensitive to the perils of generalising to the natural language case, Dummett is convinced that we can impose some version of the constraint.

So, let us assume pro tempore that some such requirement can be justified.

Dummett insists that there be a harmony between:

(1) the conditions and (2) the consequences of an assertion. Suppose that we understand by 'conditions for assertion' for \( s_k \), the verification or falsification

\textsuperscript{10} Dummett is very vague about how this generalisation is supposed to proceed. Prawitz expresses the broad intuition involved in requiring harmony between the conditions for asserting a sentence and the consequences that can be inferred from its assertion thus: 'Roughly speaking, we cannot infer other conclusions than such as must hold in view of the condition for asserting the sentence' in 'Meanings and proofs: on the conflict between Classical and Intuitionistic Logic' p. 24 in Theoria, vol. 43, 1977, pp. 2-40.

\textsuperscript{11} Reported in Prawitz, D 'Philosophical aspects of Proof Theory' p. 245, loc. cit.
conditions of \( s_k \). Does Dummett mean by 'consequences', simply the logical consequences a hearer is entitled to draw from a sentence asserted in his presence? Apparently not. For he considers the possibility of taking the meaning of a sentence as being given not by (1), its assertibility conditions but by (2) its 'consequences' in the following passage:

'Would it not be equally feasible to adopt the reverse procedure, and take the meaning of a sentence as given in terms of the consequences of uttering (asserting) it, that is, roughly, what accepting it would lead you to do that you would not otherwise have done? In the general case, this seems highly problematic, because of the often remarked fact that what difference a belief makes to your behaviour depends upon your wants'.

So Dummett apparently means by 'consequences', behavioural differences pursuant upon accepting the sentence as true. Then, the general requirement of harmony between (1) and (2) would be:

There must be harmony between the conditions that suffice to justify the assertion of a sentence \( s \) and any behavioural differences the assertion of \( s \) would produce in the speaker or hearer. \( \ldots \quad (A) \)

Take a simple example: Jones utters 'I'm in pain', acting all the while as if he's on top of the world. Because there is no harmony between his assertion and his actions, we have good reason to suspect that he might misunderstand the meaning of 'pain', a suspicion, let us suppose, borne out by further behavioural evidence; upon those grounds, we can criticise his use of 'pain'.

As for Jones' pain-behaviour, so for his logical behaviour - we find that his use of 'or' or 'not' is in disharmony with his understanding of the atomic fragment of his language - he asserts \( p \lor \lnot p \) even where \( p \) is non-ED, or he

---

12 Comments on Prawitz' p. 17 in von Wright, G.H (ed.) Logic and Philosophy, Martinus Nijhoff, The Hague, pp. 11–18. Dummett goes on to say that this might not prove an insuperable obstacle, but gives no indication as to how it might be overcome. Prawitz, on the other hand, does seem to identify Dummett's 'consequences' with logical consequences. Cf. Frege p. 453: 'Here "consequences" must be taken to include both the inferential powers of the statement and anything that counts as acting on the truth of the statement'.

13 Of course this is only one hypothesis amongst a plethora of others.
assents to $\neg \neg p \rightarrow p$. We thus have precisely the same grounds for criticising his use of 'or' and 'not' as we had for criticising his use of 'pain'.

The obvious objection to this is that if 'pain' means 'happy' in Jones' idiolect, his cheerful behaviour is not in disharmony with his use of that term - to the contrary; similarly if he means classical negation by 'not', then, in his idiolect 'not' is in harmony with his observable logical practice. Moreover this is all simply more grist for the holist's mill: any evidence supporting the thesis that Jones meant 'happy' by 'pain' is equally evidence that he wished to demonstrate that he rejoiced in his temporary affliction or that he believes that pain subsides if one acts as if one is untroubled by it or... If, then, what it is for an agent to mean something by his words and to have certain propositional attitudes is just for him to act in such a way as to satisfy an interpretative theory, if there can be no sense to the question 'What does Jones really mean by "pain" or "or" or "not"?' beyond the query as to what meaning an interpretative meaning theory would impute to these words in the context of making the best overall sense of his actions, the revisionist's project is doomed unless he can show that the classical meanings an interpreter might read into Jones' words are straightforwardly incoherent. The problem is that while Dummett does indeed have a powerful argument purporting to establish precisely this conclusion, it is unavailable to him to use against the holist who cites Jones' use of classical modes of reasoning as evidence that Jones does in fact possess the very conception Dummett declares to be incoherent. For to do so would be question-begging on Dummett's part.
§4.3: MODEL THEORY, PROOF THEORY AND THE MEANINGS OF THE LOGICAL CONSTANTS.

Dummett, as we've seen, rejects modest theories of meaning on the grounds that they fail to explicate the logic of the object language. How then will a 'full-blooded' theory of meaning fulfil this task? There are two ways in which this responsibility might be discharged - proof-theoretically or model-theoretically. These differing methods correspond to two different ways in which the relation of logical consequence can be characterised - viz. syntactically or semantically.

On a Tarskian semantic approach, one analyses what it is for an arbitrary sentence to be true-in-a-model. A sentence A is then said to be a semantic consequence of a set of sentences $\Delta$ if A is true in every model in which all members of $\Delta$ are true.

Although Prawitz has, as we shall see, expressed some scepticism on this point, most logicians would agree that Tarski's method does explicate what it is for the truth of all members of $\Delta$ to necessitate the truth of A - i.e. for A to be a logical consequence of $\Delta$.

On a proof-theoretic approach, one syntactically specifies effective rules for forming and transforming wffs of a formal language L and a sentence A is then said to be a syntactic consequence of $\Delta$ if one can move from $\Delta$ to A using only the specified rules. The syntactic rules are, naturally, selected with an eye to preserving the semantic consequence relation and, if these rules can be proved to be sound with respect to a given semantics, it will appear plausible that we have succeeded in characterising proof-
theoretically that very same relation of logical consequence which we could have characterised semantically within the given model theory.

Now I am taking it that there is one relation of logical consequence which we can characterise either proof-theoretically or model-theoretically, even if for the logician's purposes it may be necessary to bifurcate this relation into separate relations of 'syntactic consequence' and 'semantic consequence'.

We should therefore distinguish the relation of logical consequence, properly so called, from the relation of formal derivability (symbolised by ' |- ') which holds as between uninterpreted wffs of a formal language. Just because the wffs are uninterpreted, the notion of one wff following of necessity from another has no application there.

In order to characterise the deducibility relation proof-theoretically, one must be able to determine whether a given string of symbols is a formal derivation without adverting to its interpretation. Even so, the deducibility relation between sentences, although syntactically specified, must still be conceived as holding between meaningful statements. This point will be important in what follows.

As noted previously, intuitionists have generally agreed with Gentzen in regarding the introduction rules in a ND system as conferring meanings on the constants, since the introductions of the constants run parallel to Heyting's informal semantics for intuitionistic logic. We also noted that both Dummett and Prawitz argue that in order to explicate the meanings of the connectives in the Heyting explanations, we need to replace the informal

---

1 I will thus take ' | - ' to mean: ' is derivable from in '. For a clear discussion of the difference between formal derivability and proof-theoretic characterisations of deducibility, see Jonathon Lear's *Aristotle and Logical Theory*, Cambridge University Press, Cambridge, 1980, ch1.
notion of proof by a restricted type of proof. The reason given at §4.1 was that we could better account for the actual practice of intuitionists by doing so.

But Dummett gives a more compelling reason for choosing to restrict the notion of 'proof' in the Heyting clauses to a special type of proof, 'canonical proof', rather than expand it to include EM for finding canonical proofs at Elements of Intuitionism pp. 392-3. It is this: unless we do restrict the notion of 'proof' in these explanations, the explanations of the meanings of \( \rightarrow \) and \( \forall \) will be circular. For without certain restrictions, we might use the constant in question as a major premise in an application of the elimination rule for that constant in the constructions alleged to prove a statement with the constant as main operator (its introduction). Thus, in explaining the meaning of \( \rightarrow \) we cannot allow unrestricted use of Modus Ponens, the elimination rule for \( \rightarrow \) since the explanation of the meaning of the constant would be vacuous if we did. That explanation held that we have a proof of \( A \rightarrow B \) iff we have an EM for transforming any proof of \( A \) into a proof of \( B \).

Were we to allow unrestricted use of Modus Ponens in the context of that explanation, Dummett contends, we could admit any construction we liked as a proof of \( A \rightarrow B \) and it would still be true that we had an EM for transforming any proof of \( A \) into a proof of \( B \). - viz: add the proof of \( A \rightarrow B \) to the proof of \( A \) and infer \( B \) by Modus Ponens. So we cannot admit Modus Ponens as a major premise in the EM for transforming proofs of \( A \) into proofs of \( B \). In order to explain the meanings of \( \rightarrow \) and \( \forall \) in the Heyting clauses

\[ Vx.Fx \]

Dummett gives a similar argument for a restricted type of proof being used in the explanation of the meaning of \( \forall \) at Elements of Intuitionism pp.393-4. It is, again, that we cannot allow \( \forall x.Fx \) as the major premise of an application of Universal Instantiation if we're to avoid circularity.
for the connectives, we must therefore replace each occurrence of 'proof' by 'canonical proof'.

I am not convinced by this argument of Dummett's for I find it hard to see how the circularity in the argument for B above derives from an unrestricted use of Modus Ponens. Suppose \( \theta \) is a construction purporting to prove \( A \rightarrow B \) - it might be the simple instruction 'Assert "A \rightarrow B" now!'. Now suppose we have a proof of \( A \) (which we can denote by \( \nu \)). Dummett's claim is that by allowing MP as an inference rule in the EM for transforming proofs of \( A \) into proofs of \( B \), we can guarantee that we have an EM for transforming proofs of \( A \) into proofs of \( B \), a fortiori that we have a proof of \( A \rightarrow B \). So let us test this claim. We have

\[
\begin{array}{c}
\text{A} \\
\hline
\nu \\
\text{A} \rightarrow \text{B} \\
\hline
\text{B}
\end{array}
\quad (\rightarrow \text{B})
\]

Clearly the inference from \( A \) and \( A \rightarrow B \) to \( B \) will only be valid if \( \theta \) is itself a proof of \( A \rightarrow B \). But what grounds do we have for thinking that \( \theta \) is a proof of \( A \rightarrow B \)? None - and in the case at hand where \( \theta \) is simply the instruction to assert 'A \rightarrow B' now, we have every reason for believing that \( \theta \) is not a proof. So the fact that the derivation of \( B \) from premises \( A \) and \( A \rightarrow B \) fails to qualify as an EM for transforming a proof of \( A \) into a proof of \( B \) has nothing to do with the unrestricted use of MP - it is rather a direct result of the fact that \( \theta \) is not a genuine proof of \( A \rightarrow B \). The circularity in the argument derives not from the fact that we are allowing unrestricted use of MP but rather from the fact that we can only claim that \( \theta \) is a proof of \( A \rightarrow B \) if we assume the very thing we're trying to justify - viz. that this construction is one that proves \( A \rightarrow B \).

\[3\] I argue in §4.5 that there is nothing objectionably circular in using MP (or any other) inferences in the course of constructing an argument for the justification of deduction.
Dummett's argument for the canonical proof/demonstration distinction thus fails. So it seems that we should modify the informal explanations of the meanings of the intuitionistic logical constants to allow, for example, a disjunction to be asserted when we merely have in our possession an EM for proving either disjunct.

There might yet be other grounds on which the CP/demonstration distinction can be made out but I shall not investigate whether this is so here.

Dag Prawitz has shown how a recursive specification of the concept of canonical proof can be obtained for wffs of first order logic. Call a system of first order logic S an atomic system if it consists of:

1. A set of individual, operational and predicate constants.
2. A set of inference rules which have atomic wffs as premises and conclusions.

Then, we can provide a precise characterisation of canonical proof for wffs of S by adding to the modified Heyting recursion clauses the clause:

A construction C is a canonical proof of an atomic wff A iff it is a proof of A in an atomic system S.

Now suppose that we wished to encapsulate this notion of canonical proof in the intuitionistic ND system I. The idea would be to depict canonical proof as comprised of applications of introduction rules alone, since the consequences of an application of an introduction rule for a constant fulfill the conditions stated on the left hand side of the appropriate Heyting biconditional just when its premises meet the conditions stated on the right hand side.

---

4 At 'Ideas and Results in proof theory' loc. cit. pp. 242-3.
Consider (&I) by way of illustration: A canonical proof of (A & B) will have the form:

\[ D_1 \quad D_2 \]

\[ A \quad B \]

\[ (A & B) \]

with \( D_1 \) and \( D_2 \) canonical proofs of \( A \) and \( B \) respectively. So the form of the canonical proof (CP) of \( (A & B) \) is:

\[ D_1 \]

\[ D_2 \]

\[ (A & B) \]

Label this proof \( D_3 \). Then it is clear that \( D_3 \) will be a CP of \( (A & B) \) iff it is a CP of \( A \) and a CP of \( B \). The Heyting clause for '\&' is thus satisfied by \( (&I) \) in the manner indicated above.

But whilst CPs of \( (A & B) \), \( (A \lor B) \), \( \exists x.Fx \) can all plausibly be thought of as comprised of applications of \( (&I) \), \( (vI) \), \( (E) \) alone, this is not the case for the CPs of either \( A \rightarrow B \) or \( \forall x.Fx \). For even though \( (\rightarrow I) \) and \( (\forall I) \) satisfy the Heyting clauses for '→' and '∀' respectively, there are no constraints in \( I \) on the conditions that suffice to infer either \( A \rightarrow B \) or \( \forall x.Fx \) comparable to those on CPs of those wffs in the Heyting interpretation. All that we need in order to be able to infer \( (A \rightarrow B) \) in \( I \) is a proof of \( B \) from \( A \) as hypothesis - no restrictions are placed upon the type of proof involved. So, it is an open possibility that a proof of \( B \) from \( A \) will consist of applications of \( (\rightarrow E) \) as well as \( (\rightarrow I) \).

This is a barrier to our project of encapsulating the requisite notion of canonical proof in \( I \). For whilst \( (\rightarrow I) \) assuredly does satisfy the informal Heyting recursion clause for '→', it does not, initially, satisfy the given Heyting clause after it has been suitably modified in terms of canonical
proof. Yet only the 'canonical' clauses are held by Dummett and Prawitz to explicate the meanings of their associated constants. The meanings of \( \rightarrow \) and of \( \forall \) as represented by (→I) and (∀I) in I are therefore weaker and more general than those meanings conveyed by their canonical Heyting clauses. Moreover, (→I) and (→E) remain in harmony even when the premise of (→I) is permitted to be any deduction whatsoever of the consequent from the antecedent.

Prawitz has suggested a way around this problem. A closed derivation D in a system S of a wff \( \phi \) is one which has no open assumptions and no parameters that are not proper. We stipulate that a CP is to be a construction whose initial premises are atomic sentences, whose final conclusion is a sentence and for which every compound sentence in a closed derivation is deduced from its immediate premises by an introduction rule. Using this approach, we can encapsulate the notion of canonical proof for first order logic within an intuitionistic ND system. Yet, as Dummett stresses at William James Lectures, 7:26–27, it is important to allow \( (A \rightarrow B) \) in the general case to be assertible just when we have any deduction whatever of the consequent from the antecedent if we're to remain faithful to our intuitive understanding of 'if' in ordinary discourse. So, there remains a doubt as to the applicability of Prawitz's strategy outside of the contexts of constructive mathematics and logic.

---

5 Pari passu for \( \forall \). Recall that the informal clause for \( \rightarrow \) says: 'Construction C is a proof of \( A \rightarrow B \) iff C is a construction that converts each proof of A into a proof of B'. We obtain the requisite 'canonical' clauses by simply replacing each occurrence of 'proof' with 'canonical proof'.

6 Again the same result holds for (∀I) and (∀E).

7 Prawitz discusses this at 'Ideas and Results in proof theory', loc. cit. App.A 284–90.
just because they cannot be conclusively verified, it cannot be correct to take the meaning of universal or negative existential empirical statements such as 'nothing travels faster than the speed of light' as given by their conclusive verification conditions. It is far more plausible to regard the meanings of such statements as fixed not by the grounds on which we assert them but by the consequences that flow from an assertion of them. In the ND case, this would be to take the Elimination Rules ($\to E$) and ($\forall E$) as fixing the general meaning that '$\to$' and '$\forall$' carry in all contexts. On this approach, $H$ knows what $S$ is saying in asserting a conditional or a universally quantified statement just in case he knows what he, $H$, is entitled to infer from $S$'s utterance of statements of this form - he does not have to know the type of evidence $S$ had in his possession which caused him to issue such an assertion (indeed in an actual case this might be undiscoverable), but he must understand that Modus Ponens entitles him to conclude that $S$ commits himself to $q$ insofar as he commits himself to $p$ and $p \to q$, that Universal Elimination entitles him to impute the belief that $F_a$ to $S$ who asserts $\forall x . F_x$ etc.

Prawitz has recently developed a more general proof-theoretic approach to the notion of logical consequence and the meanings of the connectives\(^\text{8}\) in which the fundamental explanatory role is played not by the notion of canonical proof but by the more general notion of a canonical argument. An argument here is simply an arbitrary collection of linked inferences.

preferably arranged in tree form to make the links explicit. As before, the approach is based upon an assertibility conditional theory of meaning. But the condition for asserting a sentence is not to know a proof of it, but rather to know a closed valid argument for it. To know that B can be correctly inferred from the assumption of A is to know that we can correctly assert B in just those situations in which we can correctly assert A. So, if we have an account of the conditions under which A is assertible, we will eo ipso know when an inference involving A is valid on such a theory of meaning.

Prawitz wants to make precise Gentzen’s idea that the introduction rules are self-justifying—they are self-justifying in the sense that by stipulating what is to count as a proof of a certain conclusion, the introduction rules thereby determine the meaning of that conclusion. The elimination rules are then justified in the light of those meanings.

Now a closed argument is one for which all its assumptions and parameters are bound by some inference; otherwise it is said to be open. Open arguments are schemas for obtaining closed arguments in this sense: one can obtain from them a closed argument by replacing free occurrences of parameters by terms and free assumptions by closed arguments.

To know a closed argument for a sentence p is not to have a proof of it— to have that the closed argument must be a valid one. But how do we decide when a closed argument is valid? Prawitz makes two assumptions in order to answer this question:

---

9 More precisely, an argument can be defined as a tree of wffs with a specification of (i) which wffs are assumptions; (ii) at what places the assumptions are bound (if bound at all); (iii) at what places the parameters are bound (if bound at all), a bound parameter being debarred from appearing in a free assumption.

10 In the case of mathematics, a closed valid argument just will be a proof.

11 ‘Approaches to Logical Consequence’ loc. cit. p. 162.
(1) For each form of sentence occurring in the argument, there exists an
introduction rule for that sentence;

(2) For each inference in the argument that is not an application of an
introduction rule, there is associated with it a justifying operation which
transforms certain arguments to another argument for the same wff depending on
not more assumptions than the transformed one.

The idea is that by stating the introduction rules, one thereby determines
the canonical forms of arguments for the sentences in question, and in so
doing the meanings of those sentences is determined in terms of the
meanings of their constituents. So, an argument which has a certain
compound sentence as its conclusion, should have the relevant introduction
rule as its last inference if it is to be in canonical form.

Canonical arguments, like canonical proofs, are correct forms of argument
(proof) for the sentence in question, which are such that any correct
closed argument (proof) for the sentence in question could always
be given in these forms12.

Thus, if we ask why (&I) is a correct form of argument, the only possible
answer is that it is part of the meaning of 'and' that a conjunction is proved
by proving both its conjuncts. (The introduction rules are self-justifying). But
what of closed arguments involving elimination inferences - when are they
valid?

Prawitz thinks that certain justifying procedures can be assigned to these
inferences which transform a given closed argument containing them into a
valid canonical argument:

12 'Approaches to Logical Consequence' loc. cit. p.163.
(i) A **canonical argument** is valid iff its **immediate sub-arguments** are valid\(^{13}\).

(ii) An argument \(A\) will be valid just when it is valid with respect to some justifying procedure \(\emptyset\).

(iii) An **inference rule** is judged **valid** just when each application of it preserves the **validity of arguments**.

**Introduction rules** are then (trivially) **valid** because being in **canonical form**, they consist of valid **immediate sub-arguments**.

**Elimination rules** are **valid** iff one has a justifying procedure \(\sigma\) such that if \(A\) is an argument whose last inference is \(R\) and whose **immediate sub-arguments** are valid with respect to justifying procedures \(\emptyset\), then \(A\) is also valid with respect to \(\emptyset\) \(\cup\) \((\sigma)\).

Let us take a simple example to illustrate Prawitz's idea:

Consider Modus Ponens once more. Let \(A\) be an argument whose last inference is \((\rightarrow \emptyset)\). \(A\) will be of the form \(\emptyset \rightarrow \emptyset \emptyset\)

\[\emptyset \quad \emptyset \quad \emptyset\]

Suppose that \(A\) consists of sub-arguments \(A_1, A_2\) for \(\emptyset \rightarrow \emptyset\), \(\emptyset\) respectively which are valid with respect to justifying procedures \(\mu_1, \mu_2\). Then we can use \(\mu_1\) and \(\mu_2\) to transform \(A_1, A_2\) into arguments in canonical form \(A_{1e}, A_{2e}\), respectively. Thence we can use \(A_{1e}\) to find an argument \(A_3\) in canonical form establishing \(\emptyset\) from \(\emptyset\) as hypothesis. But then we have a justifying procedure \(\pi\) for transforming our original argument for \(\emptyset\) (i.e. \(A\)) into one in canonical form: viz. use \(A_{2e}\) to provide a canonical argument for \(\emptyset\); then use \(A_3\) to argue canonically for \(\emptyset\) from \(\emptyset\) as hypothesis. The resultant argument:

\[^{13}\] 'Immediate sub-arguments' are simply arguments for the premises of the last inference of the argument.
A sentence $p$ is said to be a logical consequence of a finite set of sentences $\Delta$ when there is a logically valid argument for $p$ from $\Delta$ (i.e., all its free assumptions belong to $\Delta$).  

Now the notion of a canonical argument for a sentence seems to hold out more promise of application to empirical statements than does the notion of canonical proof, which seems more or less restricted to mathematical and logical discourse. Moreover, the notion is not restricted to first order theories.

But is it really plausible to credit speakers who grasp the meaning of a statement with implicit knowledge of a closed valid argument for it? The Anti-Realist's claim presumably would not be that speakers could actually produce such an argument for the statement (even under epistemically ideal circumstances), but rather that speakers are disposed upon reflection

---


15 loc. cit. p. 166.
to acknowledge such an argument as correct when it is presented to them\textsuperscript{16}.

Prawitz's assumption that if a complex sentence can be asserted at all, it must be derivable from atomic sentences via some valid argument seems to me to be highly questionable however\textsuperscript{17} (unless we permit, as Prawitz clearly does not, classically valid arguments). Thus I might select one ball bearing from a box of one hundred, christen it "Alphonse", return it to the box without looking or noting its position in any way, and, though quite unable to distinguish Alphonse from the rest, still be entitled to assert 'one of these ball bearings is called "Alphonse"'. Or, to use one of Dummett's favoured examples, I might observe that I have crossed every Konigsberg bridge without observing that, as I must have, I crossed some bridge twice - that I crossed all the bridges is the content or my observation, I have not reached it (and now cannot do so) as a conclusion from discrete observations of my crossing each individual bridge (and one of these twice)\textsuperscript{18}.

Now Prawitz sees his proof-theoretic approach to validity as a genuine competitor to a Tarskian model-theoretic approach, which he contends is quite unexplanatory.

Tarski analysed logical consequence along the following lines:

\textsuperscript{16} This point highlights an important question which has not received the attention it deserves: viz. what types of behaviour on the part of speakers from the Anti-Realist's standpoint manifests their semantic knowledge? It is quite inadequate, since excessively programmatic, to reply: The activities of criticising and justifying assertions. For not all such criticisms manifest one's understanding - only criticisms directed at the content of the assertion, and what behavioural markers indicate that a criticism is so directed? Prawitz's proposal has the great merit of making this programmatic response contentful.

\textsuperscript{17} I present what I take to be a strong counter-example to it at S4.5 in which a disjunction appears as both true and assertible even though we cannot in principle tell which disjunct holds.

\textsuperscript{18} Dummett uses this example to reject Prawitz's assumption at WJL7.
Let \( A(c_1, c_2, ..., c_n) \) and \( B(c_1, c_2, ..., c_n) \) be sentences with \( c_1, c_2, ..., c_n \) standing for the non-logical constants in \( A \) and \( B \); let \( A(\nu_1, \nu_2, ..., \nu_n) \) and \( B(\nu_1, \nu_2, ..., \nu_n) \) be open WFFs corresponding to \( A(c_1, c_2, ..., c_n) \) and \( B(c_1, c_2, ..., c_n) \) respectively which are obtained by replacing the constants \( c_i \) by variables \( \nu_i \). Then:

\[
B(c_1, c_2, ..., c_n) \text{ is said to be a logical consequence of } A(c_1, c_2, ..., c_n) \text{ iff: }
\]

Every model satisfying \( A(\nu_1, \nu_2, ..., \nu_n) \) also satisfies \( B(\nu_1, \nu_2, ..., \nu_n) \). \( ... (V) \)

Now whilst Prawitz acknowledges the correctness of Tarski's criticisms of the view that logical consequence and derivability in a formal system are one and the same, and maintains that 'almost regardless of foundational viewpoint, the Tarskian... account must be accepted as far as it goes'\(^{19} \), his complaint is 'but it does not go very far'.\(^{20} \) The reason he gives is that once we make the domains of the quantifiers explicit, rewriting \( (V) \) as:

\[
(\forall \nu_1 \in D_1) \ldots (\forall \nu_n \in D_n)((A(\nu_1 ... \nu_n) \rightarrow B(\nu_1 ... \nu_n)) \text{ is true} \quad ... (V*)
\]

we see that Tarski's explication simply amounts to the claim that a sentence \( S \) is logically true if the logical sentence obtained by taking the universal closure of the WFF arising from \( S \) by replacing non-logical constants by variables is \textit{true} irrespective of the choice of independent domains. But this just shifts the original question, Prawitz contends, to the question of what it means to call a logical sentence true - i.e. to the meaning of \((V*)\)\(^{21} \) and, whilst Tarski has an analysis of truth, this analysis does not distinguish logical truth from factual truth. Prawitz concludes:

Tarski gives no analysis of the necessity of logical truth nor to answering the question of the ground for a universal truth like \((V*)\) or how we can come to know even with certainty that a logical sentence is true in all domains.\(^{22} \)

---

19 'Approaches to Logical Consequence' p.154

20 loc cit

21 loc. cit. p.155

22 loc. cit. p.154
Prawitz's criticisms of Tarski are very compressed here, but even so do not seem wholly justified to me. It is true that Tarski does not answer the question of how we can come to know that a given sentence is logically true or a consequence of another sentence, nor does he set out to. Anti-realistically, this might appear as a serious failing, but from the classical standpoint it is not. The more serious accusation concerns the necessity of logical truth and of logical consequence and here it seems to me that Prawitz's criticisms are prima facie unfounded.

Logical truths are precisely truths that hold good in all structures, according to Tarski. Quantification over all structures is supposed to capture the relevant notion of logical necessity in exactly the same way as quantification over all possible worlds is supposed to capture the more general notion of necessity. Moreover, just because logical truths are true in all models and empirical truths only true in some, we do have, on a Tarskian account, a means of distinguishing logical truths from empirical truths.

At 'General Proof theory', Prawitz tries to show by means of a concrete example why the Tarskian account is inadequate:

'Whether \( \exists x \neg P(x) \) follows logically from \( \neg \forall x P(x) \) depends, according to Tarski's definition, on whether \( \exists x \neg P(x) \) is true in each model \((D, S)\) in which \( \neg \forall x P(x) \) is true ... whether there is an element \( e \) in domain \( D \) that does not belong to \( S \) whenever it is not the case that every \( e \) in \( D \) belongs to \( S \) - i.e. we are essentially back to the question of whether \( \exists x \neg P(x) \) follows from \( \neg \forall x P(x) \).'

Prawitz is contending, in other words, that in order to establish that it is not case that everything is \( P \) entails that something is not \( P \), we must use an argument in the metalanguage that uses that very inference.

---

23 We shall shortly consider a much clearer version of this criticism of Tarski which Prawitz formulates at 'Towards a foundation of a general proof theory' in Suppes, P et al., (eds.) Logic, Methodology and Philosophy of Science IV, North-Holland, Amsterdam, 1973, pp.225-250.
A typical model-theoretic explanation of the (classical) validity of the
inference \( \neg \forall x.P(x) \rightarrow \exists x.\neg P(x) \) seems to me to be a little more informative
than Prawitz allows. One would normally argue for the validity of the
inference as follows:

'In any model, \( \exists x.\neg P(x) \) is either true or not true.
(a) Consider those models in which it is true; then if \( \exists x.\neg P(x) \) is true, a fortiori \( \neg \forall x.P(x) \)
\( \rightarrow \exists x.\neg P(x) \) must also be true.
(b) Consider those models in which it is not true; since \( \neg P(x) \) is false of every \( d \in D \), \( P(x) \)
is true of every \( d \in D \). Hence, \( \neg \forall x.P(x) \) is false and thus \( \neg \forall x.P(x) \rightarrow \exists x.\neg P(x) \) must be true.'

This is a non-circular argument in the semantic metalanguage for the
validity of a wff (or an inference). But Prawitz would reply that the
argument simply reduces the questioned implication from 'it is not the case
that everything is \( P \)' to 'something is not \( P \)' to another implication:viz. - from
'not \( P \) is false of everything' to 'it is false that something is not \( P \) - which
while not the very same implication as the original is sufficiently close to
that implication\(^{24}\) to make the model-theoretic 'explanation' quite
unilluminating: a Tarskian demonstration of the validity of that inference
only goes through on the assumption that other laws hold in the
metalanguage, which assumption is never justified. It seems to me that we
must concede that Prawitz is essentially right about this.

In contrast to Prawitz, Dummett does not see a proof-theoretic account of the
meanings of the connectives as a genuine competitor to a model-theoretic
account. On the contrary, even when we choose a suitably restricted class of
inference rules to perspicuously represent the meanings of the logical
constants, these rules can only succeed in doing so if we already have an
implicit grasp of how the semantic value of a sentence depends upon the

\(^{24}\) Given what Dummett calls the Negation Principle - \( p \) is false iff it is not the case that \( p \) is true
- the two are interderivable.
semantic values of its parts. Proof-theoretic characterisations of the meanings of the constants are for Dummett *supervenient upon* a characterisation provided by a semantic theory. To grasp the meaning of a connective such as 'or' simply *is* to have a tacit understanding of how the semantic value of complex sentences involving it are determined by the semantic values of their parts.\(^\text{25}\)

Dummett lays down some fairly stringent conditions that a set of inference rules must satisfy if it is to provide a perspicuous representation of the meanings of the connectives. But, even if we take the meanings of the constants to be determined by the stipulation that these inference rules are to hold, Dummett contends, we still do not have a genuine *alternative to* the above semantic conception of their meaning.\(^\text{26}\) For each inference rule in the set has to be understood as contributing to the determination of the meaning of the constant by displaying either a necessary condition for the truth of a complex sentence involving it (in the case of an elimination rule), or a sufficient condition for the truth of such a sentence (in the case of an introduction rule). A complete set of (suitably constrained) introduction rules, by displaying all possible sufficient conditions for the truth of the complex sentence, or a complete set of (suitably constrained) elimination rules by displaying all possible necessary conditions for the truth of that sentence, gives a necessary and sufficient condition for the truth of the sentence.\(^\text{27}\) But just because the proof-theoretic characterisation yields a truth-condition for the relevant compound sentence, the precise meaning that the inference

\(^{25}\) Cf. WJ I: 15ff.

\(^{26}\) loc. cit.

\(^{27}\) Since the introduction and elimination rules are required to be in harmony, it is possible to choose either sort as explicating the meanings of the constants — which one chooses will probably depend on whether one takes the conditions for asserting a sentence or the consequences of asserting it as the central concept in one's theory of meaning. Of course, one could choose both.
rules confer upon the constants will depend upon the concept of truth we’re appealing to. On Dummett’s view, however, the concept of truth will be determined by the theory of meaning – which, to recall, is a three-tiered structure consisting of a semantic theory as base, a theory of sense correlating speakers’ grasp of truth-conditions with practical abilities, and a theory of force connecting the sense of each sentence with its actual use in discourse. Hence, since the theory of meaning is founded on a semantic theory, proof-theoretic characterisations of the meanings of the connectives (and thus of validity) must supervene upon a semantic characterisation.

Dummett’s argument in brief is that since proof-theoretic characterisations of the meanings of the connectives presuppose a grasp of the concept of truth and since the meaning theory which purports to explicate the concept of truth presupposes a grasp of the meanings of the connectives, proof-theoretic characterisations of the connectives presuppose a grasp of the meanings of the connectives.

I think this argument is open to question– for the part of the theory of meaning which is presupposed by the concept of truth on Dummett’s model – i.e. either the theory of sense alone or the theory of sense together with the theory of force – does not include the part which is responsible for explicating the meanings of the constants. Dummett’s argument would only go through if it were the same part that was required to simultaneously explicate the meanings of the constants and the concept of truth.

As it is, we have, on Dummett’s picture, a theory of meaning as comprised of a recursive specification of the meanings of the connectives as base, a theory of sense which (either by itself or possibly together with the theory of force)
explicates the concept of truth and the theory of force. The recursive semantic theory presupposes the concept of truth, the articulation of which is supplied by the theories of sense and force. The question is what form this should take - perhaps a model theory or, in its place, the introduction and/or elimination rules for the connectives.

Now we might, for Prawitz-styled reasons, choose to explicate the meanings of the connectives proof-theoretically rather than model-theoretically, the idea being that the Gentzen-styled introduction and/or elimination rules are to be justified in terms of our intuitive grasp of logical consequence without themselves supplying an analysis of logical consequence. If we did this, there would be no circle of semantic dependence - the proof theory would supply the meanings of the constants; the theory of sense (and possibly force) would specify the truth-conditions for the sentences of the object-language and these tasks are distinct. So the fact that the recursive specification of the meanings of the connectives takes truth as primitive and the theory of meaning takes the meanings of the (metalinguistic) connectives as primitive is unproblematic (it does not show that we must have a tacit grasp of the meaning of 'and', as Prior thought, before we can understand (&1), for example, or indeed of any overtly semantic notions as Dummett thinks). For each part of the total theory of meaning attends to different tasks.

Dummett imposes four main constraints on any putative meaning-revealing set of inference rules\(^\text{28}\). The rules should be individually (1) Pure; (2) Simple and collectively (3) in harmony; (4) Stable.

A Pure inference rule is one for which no more than one logical constant appears in the formulation of the rule. Modus Tollens is an impure rule. In

\(^{28}\) In *WJL6*, he mentions the first three; in 'Comments on Prawitz', loc. cit., the fourth.
order for a rule to be simple, several conditions must simultaneously be satisfied:

(i) There must be only one occurrence of the constant in a statement of the rule: 

\[(A \rightarrow B) \rightarrow A \quad \text{and} \quad \neg A \rightarrow A\]

are thus not simple.

(ii) The constant must appear either in the premises or in the conclusion but not in both; i.e. the rule must be unambiguously either an introduction or an elimination rule.

\[\neg A \quad \vdash \quad \neg A\]

\[\neg A \quad \vdash \quad \neg A\]

violates this condition.

(iii) While one or more subordinate deductions may figure among the premises, the logical constant itself must not appear in one of the hypotheses of the subordinate deductions (i.e. one of the hypotheses discharged by the inference rule). Non-Constructive Dilemma, (NCD)

\[\neg A \quad \neg A \quad B \quad B\]

violates this condition.

What is the justification for these demands? Dummett contends that unless the inference rule relates to the most general form of sentences involving the constant in question, it cannot directly determine the meanings of sentences not of the form cited in the inference rule - hence it must be pure and simple in sense (i). It must be simple in senses (ii) and (iii) because if we are giving the meaning of the constant in terms of its inferential role in a
complex statement, we cannot presuppose a grasp of that statement in specifying this inferential role.

Now these are very stringent demands to place upon individual inference rules and, unsurprisingly, strictly classical rules turn out to be either impure or not simple. However, intuitionistic negation is also ruled out since it is impure\(^{29}\), although the other intuitionistic rules are both pure and simple. As to the constraints regulating sets of meaning-explicating inferences, we have already discussed the need for harmony between introduction and elimination rules. The requirement of Stability arises as follows:

Suppose that two alternative proof-theoretic procedures for justifying forms of argument are proposed - (a) An 'upwards' justification procedure which appeals to introduction rules as giving the meanings of the connectives; and (b) A 'downwards' procedure which appeals to elimination rules as effecting this task. Suppose we start with a set of elimination rules. Then the set of introduction rules that can be justified with respect to them by a downwards procedure is well defined. Now consider this set of introduction rules - the set of elimination rules that can be justified with respect to them is similarly well defined (by an upwards procedure). The question of Stability has to do with whether we will arrive back at the set of elimination rules with which we started. If so, we call that set of elimination rules stable\(^{30}\).

Stability is not the same property as harmony because the latter is automatically guaranteed by either type of justification procedure. Dummett lays it down as a further intuitively reasonable condition that:

\(^{29}\) It would be intolerable for the Anti-Realist to simply acquiesce in this state of affairs since he must give an account of negation on pain of relinquishing the concept of falsehood for the sentences of the object-language.

\(^{30}\) We can similarly define stability for a set of introduction rules.
'no intelligible meanings can be conferred on a set of logical constants by the stipulation of an unstable set of introduction and elimination rules'. 31

No assertibility conditional theory of meaning can be provided for a language with an unstable set of introduction rules governing its logical constants; no theory of meaning based on the consequences of an assertion can be provided for a language with an unstable set of elimination rules.

Now I do not wish to investigate the plausibility of Dummett's requirements here32. Suffice it to say that the need for some such constraints upon allegedly meaning-canonical inference rules seems a genuine one wherever we are dealing with a constructivist semantics which is itself based on the central concepts of inference and proof. For here, the threat of impredicativity is ever-present. Whether these constraints can reasonably be imposed upon proof-theoretic approaches to the meanings of the classical connectives seems far less certain.

Dummett regards a set of pure and simple inference rules as displaying the general form of the meanings that the operators carry in all contexts33.

Further, specific introduction rules pertain to specific contexts - e.g. whilst an intuitionistic understanding of canonical proofs (hence meaning) of \( \forall x.Fx \) is governed by the demand that \( (VB) \) be valid34, the means of establishing a universal statement may vary from domain to domain - in the particular domain of natural numbers, for example, a recognition of induction as a (in fact the only) method of proving \( \forall x.Fx \) depends upon an understanding of the construction of that domain. This is just symptomatic, according to

32 We will take up the question of whether the requirement of harmony or the more specific requirement of conservative extension can do the work that Dummett wishes it to in S4.5.
33 WJL 7:28.
34 This seems to support the need for a 'mixed' theory, specifying both assertibility conditions and deniability conditions for the sentences of the language.
Dummett, of the general point that inference rules only fix the meanings of the constants against an extensive background of semantic principles governing the language. The inference rules governing 'or' (e.g. \(\land I\) and \(\land E\)) neither require us to nor forbid us from assuming that LEM holds in the object language. But while both intuitionist and classicist agree that the truth of \(A\) and the truth of \(B\) exhaust the possible sufficient conditions for the truth of '\(A\) or \(B\)', the Realist understands truth as something that attaches to the sentence independently of our capacity to establish it, whilst the intuitionist identifies truth with our capacity to establish it. A conventionalist will object that the differing interpretations of 'or' surface in the additional inference rules that each takes to hold - e.g. the Realist accepts LEM, the intuitionist doesn't. Dummett agrees. He denies only that it follows from this that the meanings of the logical constants are completely given by specifying the totality of logical laws governing them. We can choose certain restricted types of inferential rules to fix the meanings of the connectives they govern, but what meaning they can be taken to impart to those constants depends upon the concept of truth explicated in the theory of meaning.
§4.4: ANTI-REALIST THEORIES OF MEANING

A verificationist theory of meaning takes the meaning of a statement to be given by the conditions that justify its assertion. A falsificationist theory takes those consequences of a statement that would reveal its assertion to have been incorrect as constitutive of its meaning. I believe that there are good reasons for thinking that the meanings of the logical constants, at least within the context of empirical discourse, cannot be captured in terms of a central concept of verification or falsification. The meanings of the connectives must, I will argue, be explicated in terms of the notion of truth. Thus, given that the Anti-Realist rejects verification-transcendent conceptions of truth (falsehood) and, as I hope to show, that the requisite non-Realist notion of truth is no simple construct out of assertibility (deniability), the Anti-Realist must provide a tenable alternative to classical truth if his semantic theory is to remain viable.

Fallibility, finitude and fortune conspire to make the verification or falsification of an empirical statement a very precarious thing. Verifiability and falsifiability are not in general enduring properties of empirical statements (as provability is of mathematical statements). For statements that can be verified, it is plausible to believe that there must be some optimal verification point (OVP) at which an assertion of that statement could be established as correct.

I borrow this notion from Bernard Williams' 'Another Time, Another Place, Another Person' in Moral Luck, Cambridge University Press, Cambridge, 1981. At least for the simple occasion sentences that we will look at, one can identify the OVP with some space-time position (x,t) or some set of such positions. In the general case, the OVP could be conceived information-theoretically as representing a node at which information is maximised [cf Williams p.167]. The rate of decay from this node might be very rapid or very gradual.
At Truth Conditions and Criteria', Crispin Wright claimed that an Anti-Realist could just take over the form of a Davidsonian truth-theory in giving a theory of meaning for a language - all that the Anti-Realist has to do is to read the invariant expression 'is true iff' connecting mentioned OL sentences with used ML sentences in the T-sentences as 'is justifiably assertible iff'.

Then, given an intuitionistic proof theory in the metalanguage, we can, by pairing mentioned OL sentences with their metalinguistic translations, state the meanings of all OL expressions via a structure-reflecting truth-theory for that language in an Anti-Realistically acceptable way. My discussion of Wright's proposal is heavily indebted to Anthony Appiah's recent discussion of it, although I believe many of his criticisms to be mistaken.

Wright's suggestion is that we read the Tarski biconditionals as:

(1) $s$ is justifiably assertible iff $p$.

Here, as in the Davidsonian original, $p$ is a translation into the metalanguage of the object language sentence designated by $s$. I will assume that the Anti-Realist has at his disposal a method of radical interpretation analogous to Davidson's. Wright is quick to point out that we cannot understand 'iff' as material equivalence if we wish $p$ to translate $s$ since:

'p cannot be synonymous with $s$ in any case where "$s$ is justifiably assertible" may be verified in circumstances falling short of a verification of "$s$".'

---

2 loc. cit. pp.239-240.
3 At 'Anti-Realism unrealised', Philosophical Quarterly 34, 1984.
4 I will represent 's is justifiably assertible' as $j(s)$. We will later require a ternary relation $i(s, A, t): 's$ is justifiably assertible for $A$ at $t$'.
5 'Any sentence comprising the right hand side of a Tarski biconditional be a translation of the quoted sentence on the left hand side' loc. cit. p.238
6 Wright does not even address this assumption. He needs to say something about the empirical means for testing whether metalinguistic sentences really do translate the object language sentences with which they are paired.
7 loc. cit. p.238. I have changed the symbols for Wright's designated object language sentence and used metalanguage sentence ($p$ and $\bar{s}$ respectively) to $s$ and $p$ respectively.
If such situations were to arise, we would have to conclude, contrary to our assumption, that p is not synonymous with s. Because he assumes that we can somehow know at an arbitrary point that our ML sentences really do translate our OL sentences, Wright concludes that this is absurd and that we must seek an Anti-Realistic reading of "iff". An alternative hypothesis in the face of such a highly probable occurrence is that p in fact is not synonymous with s. Still, there seems to be an independent reason for seeking an Anti-Realistic reading of the biconditional - viz. that the meaning of that connective as explicated in terms of classical truth is (Anti-Realistically) incoherent.

Wright provides this reading of the conditional at Dummett and revisionism, page 54.9

**COND**
A Total State of Information (TSI) justifies our assertion of 'if p then q' just in case we can recognise that its modification into a TSI justifying the assertion of p would transform it into a TSI capable of effective transformation into a TSI justifying the assertion of q.

Wright puts it forward as intuitively reasonable that:

'...to be justified in asserting that s is justifiably assertible is to be justified in making an assertoric use of any sentence synonymous with s'.10

Letting 'j' denote the operator 'to be justified in asserting that ...' and 'j' the sentential predicate 'is justifiably assertible', we can write this as:

(2) J(j(s)) iff Jp.

When we apply (2) to (1), we get:

---

8 Appiah follows him in this conclusion.

9 The version I present here is actually the one Wright suggests at 13n pp54-55 of that article.

10 loc. cit. p.239
Letting \( q \) denote the sentence in (1) (i.e. \( j(s) \) iff \( p \)), we can write (3) as:

\[
(3') \quad j(q) \text{ iff } (a) j(j(s)) \rightarrow j(p) \quad \text{and} \quad (b) j(p) \rightarrow j(j(s))
\]

If Wright's interpretation of the conditional (hence the biconditional) is correct, the assertion of (1) will be justified.

Now Appiah contends that (1) will not do as a T-sentence for sentences of a natural language because it takes no account of the dependence of such sentences on context\(^1\). But this is not at all obvious, and, it would seem, in explicating the meaning of the intuitionistic biconditional along Wright's lines, the relativization of utterance to state of information is revealed and TSI\(^a \) are already token-reflexive. It is not that (1) ignores indexicality, it is rather that the features of index and contribution of context are implicitly assumed to be understood by anyone who grasps the biconditional. So the task is to articulate these indexical features. In what follows, I will leave it open whether, in accord with the conclusions on indexicality arrived at in §1.3.2, the Anti-Realist will wish to explicate the contextual factors determining the correctness of an assertion of \( s \) either by the method of conditional assignments or by Lycan's V-function. For reasons of simplicity, I will adhere to a speaker-demonstration approach discussed and rejected at §1.3.2. The results can be recast in terms of either of the other two acceptable ways above.

\(^1\) loc. cit. p.89
There is in fact a serious problem with (1): Let s be any non-ED sentence; then \( j(s) \) says that s is justifiably assertible and since, for the Anti-Realist, it is recognisable whether s is justifiably assertible or not, the LHS of the biconditional is decidable. Yet the RHS of the biconditional, p, is a translation into the metalanguage of s and if s is non-ED, then so must p be. Hence, since we wish our T-sentences to be such that the RHS is (at the least) logically equivalent to the LHS, we have a patently incorrect result - a non-ED sentence is being held to be logically equivalent to (or, worse, mean the same as) a decidable sentence. So (1) must be incorrect.

Now Wright seems to have some awareness of this difficulty but could not, I think, have grasped its real import for he believes that it can be overcome (1) by first reading the 'iff' as an intuitionistic biconditional along the lines suggested in COND and then understanding the predicate 'is justifiably assertible' as subject to (2), as above. The intuitive response to this is that since an intuitionistic biconditional is some sort of logical equivalence relation ('p iff q' says that TSI's warranting the assertion of p are (in some sense) logically equivalent to TSI's warranting the assertion of q) and it is demonstrably wrong to classify ED and non-ED statements as

---

12 In fact, as we shall see, there is more than one problem with (1) but this is the most dire.  
13 Thus in discussing whether we ought to construe unasserted sentential constituents of a compound sentence as synonymous with statements of their own warranted assertibility, he explicitly states the difficulty above: 'We are not allowed in general to do so, for a simple reason: they are not the same statement. 'P is justifiably assertible (in the present state of information)' is, if true, always in principle decidable so; whereas there may be no notion of verification for P, and, if there is, the former may still be verified in circumstances falling short of it.' loc. cit. p.235.

Having stated the difficulty with treating unasserted atomic sentential parts of compound sentences as synonymous with or logically equivalent to 'statements of their own warranted assertibility', it seems hard to believe that he did not also realise that precisely the same difficulty arises when these sentences occur in assertions by themselves. What might have caused him to miss this point, if in fact he did so, was his concentrating in the quoted passage on the fact that such sentential parts occur unasserted.
logically equivalent, no amount of tampering with the precise interpretation of the biconditional or with the predicate ‘j’ in (1) can make it come out right. It now remains to see whether this response is the correct one.

If we were to take \( p \) as ‘it is raining’ we could start to spell out the context-dependence implicit in Wright’s ‘iff’ by re-expressing (1) as:

(4) \((\forall t) (\text{the assertion at } t \text{ of ‘it is raining’ is justified iff it is raining at } t)\)

But this is too inexplicit for it still does not tell us whether the assertion and its justification concern the same person at the same time. Thus Brother Ambrose cloistered away in his dark, silent study, though by chance realising a truth in asserting that it is raining, does so unjustifiably, whilst the less prophetic and more practically-minded Brother Maynard in uttering the same sentence as he scans the heavens does so with complete justice. As Appiah emphasises, sentences are assertible for particular speakers at particular times. So we at least need in place of (4):

(5) \((\forall t)(\forall \text{persons } A)(\text{the assertion of ‘it is raining’ by } A \text{ at } t \text{ is justified iff it is raining at } t \text{ near } A)\)

Generalising (5) to all sentences \( s \), we obtain:

(6) \((\forall t)(\forall A)(\forall s)(\text{ j(s,A,t) iff } p_{t,A})\).

[where the subscripts in \( p_{t,A} \) are meant to indicate that \( s \) is assertible (and, for occasion sentences such as ‘it is raining’, only assertible) for the same speaker \( A \) at the same time \( t \).]

So (4), (5) and (6) are meant to be analogues of Davidsonian T-sentences such as: “‘It is raining’ as uttered by \( A \) at \( t \) is true iff it is raining at \( t \) near \( A \).”

The great virtue of (4) is that, unlike its predecessor (1), it allows one to evaluate an assertion that has been made at a different time \( t \), now. Thus

---

14 Even though he stresses this fact, he seems to overlook it at a crucial point as I shall later show.
suppose A said 'It is raining' yesterday; we quantify over TSI s in our biconditional, TSI s which include the present one: A's assertion yesterday of 'It is raining' is now justified iff it was then raining near A. When we assert \( j(s) \iff p \) ', we're talking relative to the background of some TSI - 'is justifiable' is itself context-relative. So what (4) really means is:

'Every TSI is such that \( in it \) we can justifiably assert that the assertion-at-t of "It is raining" is justified iff \( in it \) we can justifiably assert that it is raining at t'.

So, feeding (5) (or (6)) into Wright's biconditional, we obtain:

(7) (5) is now assertible just in case we can recognise that:
(a) If A came to be justified in asserting at t that 'it is raining' is assertible, he'd come to be justified in asserting at t that it is raining.
and
(b) the converse of (a).

Now Appiah denies that (7), or at least his version of (7), is correct. His argument seems plausible, however, only because it exploits an ambiguity in the scope of the temporal quantifier. Thus instead of our (4), he has as his analogue to a Davidsonian T-sentence:

\( (R^*) \) At all times \( T \), 'It is raining' is assertible at \( T \) iff it is (tenselessly) raining at \( T \).

Correlatively, instead of (7), he has:

\( (3^* \, (R^*) ) \) is now assertible just in case we can recognise that
(a) if we came to be justified in asserting that 'it is raining' is (tenselessly) assertible at \( T \), we should come to be justified in asserting that it is (tenselessly) raining at \( T \); and
(b) the converse.\(^{16}\)

---

\( ^{15} \) It is surprising that Appiah should put this forward as the analogue to the Davidsonian paradigm since he represents the latter at p.90 as:

\( (R) \) At all times \( T \), 'It is raining' uttered at \( T \) is true iff it is (tenselessly) raining at \( T \).
Surely the analogue to (R) wherein 'justifiably assertible' substitutes for 'true' is not \( (R^*) \) but:

\( (R^\prime) \) At all times \( T \), 'It is raining' asserted at \( T \) is justified iff it is (tenselessly) raining at \( T \).

\( ^{16} \) loc. cit. p.90
Now Appiah naturally has no difficulty in demonstrating the inadequacy of (3\textsuperscript{a}), for, to modify his counterexample slightly, we can imagine Brother Ambrose fervently believing that his prophetic powers had worked on Saturday afternoon, finding out to his chagrin on Sunday morning that naughty novice Nigel had turned the sprinklers on the window he had been peering through, thereby revealing now this supposed confirmation of his powers to have been a sham. So (3\textsuperscript{a}) fails for at least one time, since even though Ambrose has come to be justified in asserting that 'it is raining' was assertible on Saturday afternoon, he has not thereby come to be justified in asserting that it was raining then\textsuperscript{17}.

For Appiah, the property of being justifiably assertible is a property that a statement can \textit{lose}; it is an essentially temporal property. But is this really the way an Anti-Realist sees it? Certainly not. To be sure, Wright does not say either at 'Truth Conditions and Criteria' or 'Dummett and Revisionism' whether the notion of verifiability he appeals to in order to explicate the meanings of the Anti-Realist's logical constants is a conclusive or an inconclusive one. But if, qua Anti-Realist, Wright thinks that 'j' in (1) \textit{j(p) iff p} need not be a conclusive variety of justification, he is surely wrong - we do not need much imagination to see that if 'j' means 'highly probable' or 'almost certain', (1) \textit{has} to be incorrect irrespective of the precise interpretation of the biconditional.\textsuperscript{18} So the concept of justification \textit{required} in our definitions so far is \textit{conclusive} justification.

\textsuperscript{17} loc. cit.

\textsuperscript{18} In point of fact there is some evidence to suggest that Wright might have thought otherwise - witness 'Dummett and Revisionism' p.50: '... in contrast with proofs, the assertion-grounds of contingent statements tend to supply a defeasible warrant for the assertion of those statements and thus cannot be construed as truth-conditions'
Now let us check whether the 'counterexample' above is still a counterexample to the (correct) (7) after disambiguating the meaning of 'j'. Was there ever a time at which Ambrose’s assertion of 'It is raining' was justified? If there was, it was on Saturday afternoon. But the Anti-Realist will simply deny that Ambrose’s assertion at that time was (conclusively) justified - the evidence that it was raining then was only inconclusive and inconclusive evidence for s is not sufficient by the lights of (1) to justify the assertion of s. Appiah will no doubt respond that inconclusive evidence is all we ever get for empirical p - in the general case, this is no doubt true (though it is surely mistaken for decidable empirical sentences such as 'It is raining'), but this is a different point, one to do with how the Anti-Realist proposes to address the question of defeasibility on a semantics which takes conclusive verification as its central concept; it is not as such a demonstration that there is anything wrong with the definitions in (1), (4), (5), (6) and (7) properly construed.

Now we can formulate an argument against the acceptability of Wright’s account of the biconditional (a fortiori, the conditional) using the already noted unacceptability of (1). It proceeds as follows:

19 This does not seem implausible to me. Would it have been rational for Maynard, on the strength of Ambrose’s claim, to form the belief that it was raining? Not if he knows anything about Ambrose’s penchant for evidenceless assertions. And even if he doesn’t and just sees him looking out and actually checking (for once), he will surely want to know that Ambrose has an unobstructed view, that there are rain clouds in the sky etc, all of which are eminently decidable matters. Appiah thinks that a simple example such as this (in his example, I am duped into believing that it is raining by seeing drops of water caused by a window cleaner fall upon my window) suffices to refute Wright’s proposal at (1) once we ‘attend to temporal quantifiers’ (loc. cit. p. 103). The irony is that it is only because he incorrectly interprets the scope of the temporal quantifiers in the first place that his supposed counterexample can even resemble a difficulty for Wright.
The definition (i) - j(s) iff p - cannot be correct because for non-ED s, j(s) is ED and p non-ED. So, prior to analysis of the biconditional, we can know (i) cannot be right.

Now initially by 'iff' we meant (intuitionistic) logical equivalence and we already know that (i) is incorrect, the basic reason for the failure of (1) being that on the LHS of the biconditional we have a claim mentioning the epistemic state of a speaker (about when the evidence he has available to him for s conclusively justifies s) and on the RHS we have a description of a state of affairs (even if one recognizable in principle) which makes no mention of epistemic states, a state of affairs of which the speaker may be quite unaware.

Whichever we require, a la Davidson, that the RHS translate the LHS of the intuitionistic T-sentence, the unsatisfactory situation in (ii) will persist.

Now if we feed into (1) Wright's account of logical equivalence (as developed in COND and (2)), we eventually obtain, after making temporal and other indexical parameters explicit in order to articulate the meaning of "iff", (7).

But so long as by 'justifiably assertible' we understand conclusive justification and so long as we read the temporal quantifiers correctly, (7) seems right.

Yet if (7) just is (1) fully explicated and (1) is incorrect and (7) correct, the only conclusion that we can draw is this: Wright's account of the biconditional cannot be correct.

A fortiori, Wright's account of the conditional of (1) cannot be correct and neither is his suggestion that an Anti-Realist can simply take over Davidson's truth-theoretic approach to meaning with the notion of conclusive justification substituting for truth.

The upshot of the above argument is that the Anti-Realist has two broad options:

(A) To reject or seriously modify the Davidsonian paradigm in which the RHS of the biconditional translates the LHS in the T-sentences of a truth-theory and develop an independent assertibility-conditional theory of meaning.

(B) To retain the Davidsonian paradigm of translation but develop an Anti-Realist conception of truth.

I will contend that he should choose option (B).

Sooner or later, the Anti-Realist must face the fact that most types of empirical statement are not conclusively verifiable. This is a point which
both Dummett and Wright acknowledge. It is unquestionable that we do assert empirical statements on the basis of inconclusive evidence. Lest this be held to be another of our practices that ought to be revised, we can note that for universally quantified empirical statements for example or indicative or subjunctive conditionals, it is quite obscure how a conclusive verification could possibly be achieved - how could we conclusively verify 'All men are mortal', for example? So the 'conclusive verificationist' faces a challenge from his 'inconclusive' counterpart similar to the one which the Realist faces from him: given that the practice that we actually acquire and engage in is to issue and accept certain assertions on the basis of inconclusive evidence, how can we possibly justify ascribing to speakers a grasp of conclusive verification conditions for such assertions? (Grasp of conclusive verification conditions is unverifiable). If we are forced to account for the meaning of such statements in terms of conclusive verification, we will be at a loss to explain our accepted grounds for making such assertions or the accepted consequences pursuant on their utterance.

In accord with (A), the Anti-Realist might wish to keep to the outward form of a truth-theory, but give up the problematic constraint that p translate s in (1) j(s) iff p, or in (6) (\forall t)(\forall A)(\exists s)( j(s,A,t) iff p_{t,A}). p would then be free to advert to TSI which suit them out for justifying the assertion of p, rather than to states of affairs (albeit recognisable ones) whose obtaining A might be quite unaware of. So we could replace (6) by:

(6 *) (\forall t)(\forall A)(\exists s)( j(s,A,t) iff A has evidence for s at t).

20 Witness Wright's remark quoted at 10n and Dummett's remark at TOE xxxviii 'it is misleading to concentrate, too heavily, as I have usually done, on a form of anti-realist theory of meaning in which the meaning of a statement is given in terms of what conclusively verifies it; often such conclusive verification is not to be had'.
We can write this as: \((\forall t)(\forall A)(\forall s)(\ j(s, A, t) \iff E(s, A, t))\). Here, the evidence symbolised by 'E' need not be conclusive evidence (for we have surrendered, pro tempore, the problematic constraint that the RHS translates the LHS).

Now \((6^*)\) seems unobjectionable (since it is acceptable to Realists and Anti-Realists alike). But it only represents the simplest case—where \(p\) is atomic. How are we to construe the clause governing negation, for example? What we seek to derive is this:

\((7) \ (\forall t)(\forall A)(\forall s)(\ j('not\cdot s\cdot, A, t) \iff E(-p, A, t))\)

[where '-' is the metalinguistic negation operator]. We need an Anti-Realistically acceptable account of negation in order to derive this. Wright suggests one:

**NEQ**

A TSI ... justifies the assertion of 'not's' just in case it justifies the assertion that a TSI justifying the assertion of 's' cannot be achieved no matter how thoroughgoing an investigation is conducted\(^{21}\).

The difficulty here lies again with the intuitionistic biconditional 'just in case'. Assume NEQ is correct. Let \(q\) denote the sentence 's is justifiably assertible' which we earlier wrote as 'j(s)'. Suppose we wish to know the circumstances under which 'not\cdot q\cdot' is assertible. Then according to NEQ:

'not\cdot q\cdot' is justifiably assertible iff there is no extension of the present TSI, \(\Sigma_{10}\), in which we could come to be justified in asserting that \(\mathcal{B}\) (where \(\mathcal{B}\) is the translation of \(q\) in the metalanguage).

We can write this as:

\((8) \ j('not\cdot q\cdot') \iff -(\Sigma_{\mathcal{T}l})(\Sigma_{\mathcal{T}l} \Sigma_{10} \& \mathcal{B})\)

where '-' is the metalinguistic negation operator and '\(\Sigma_{\mathcal{T}l} \Sigma_{10}\)' is to be read:

'TSI \(\Sigma_{\mathcal{T}l}\) extends the present TSI \(\Sigma_{10}\)'.

\(^{21}\) 'Dummett & revisionism' p.50
Let us assume that 'not' q is justified in our present TSI, i.e. \( j('not' q) \) in \( \Sigma_{10} \). Suppose there exists a TSI \( \Sigma_{1k} \) in which q is justified. Then in that extended TSI, \( j(j(q)) \). Now it is possible that whilst the assertion of s is not justified in our present TSI \( \Sigma_{10} \) (due to lack of any evidence for s), a TSI \( \Sigma_{1k} \) may emerge in the future in which the assertion of s is justified (when evidence comes to light for s in \( \Sigma_{1k} \)). But this is just to say that it is consistent with our initial assumption that \( j('not' q) \) in \( \Sigma_{10} \) that there exists a future TSI \( \Sigma_{1k} \) in which \( j(j(q)) \) - i.e. \( j(q) \), which is simply to contradict the definition of negation inNEG.

Clearly, the Anti-Realist's semantic proposals are hamstrung by the lack of an acceptable account of his biconditional since the adequacy of the explanations of the meanings of each of the other connectives depends upon a clear interpretation of that constant. But negation presents acute difficulties of its own, at least on an assertibility conditional theory of meaning as Wright notes.\(^{22}\) The fundamental worry is that since neither the assertion of p nor the assertion of not-p need be warranted by the available evidence, it cannot be correct to suppose 'not-p' to be assertible just when p cannot be asserted, however we construe the 'cannot'.

Moreover, even if an adequate account of 'iff' could be supplied for NEG, there seems to be an independent worry with the interpretation of the 'cannot' in that definition. If we suppose the 'cannot' means 'logically cannot', then we must be supposing that the notion of justification being appealed to in the definition is conclusive justification. But then we face the difficulty of defeasibility above: how are we to then explain the meaning of 'not' as it

\(^{22}\) 'Truth Conditions and Criteria' p.238
occurs in statements made on the basis of ineradicably defeasible evidence such as 'Jones is not in pain' in the teeth of the 'inconclusive' verificationist's challenge to his 'conclusive' counterpart? Either that aspect of our practice must be held to stand in need of revision or it must be admitted that NEG is not general enough to account for it. But the 'cannot' cannot mean 'empirically cannot' either, since the loss or destruction of evidence for p clearly does not justify the assertion of 'not-p'. So there are good reasons for thinking that NEG is inadequate, independently of the precise interpretation of 'iff'.

Merely by knowing the assertibility conditions of p we are not assured of knowing when the assertion of its negation is warranted - knowledge of the AC's for p does not expand itself into knowledge of the deniability-conditions (DC's) for p. So, it would seem that an adequate Anti-Realist semantics must include DC's as well as AC's for all OL sentences and, of course, a plausible mapping between them.

Let us recall the intuitionistic explanation of negation. The intuitionist has a clear conception of falsity for arithmetical identities: each such statement is false if its proof-procedure does not yield $0=0$. At least for decidable p, then, the intuitionistic explanation of the meaning of '¬p' as 'a proof of p is impossible' seems intuitively clear. We can explain what it is for a proof of p to be impossible in terms of a capacity to transform any purported proof of p into a proof of a primitive absurdity such as $0=1$. Where p is not decidable,

---

23 Although there is a possible impredicativity in this definition. For, the definition seems to require the ability to recognise of any construction C purporting to prove p that it exclude any possible construction proving p. But, if with the intuitionist, we deny that we have a fixed range of possible proofs of p, this does seem to involve an impredicativity.
we have to explain the meaning of \( \neg p \) by specifying what counts as a refutation of \( p \). Then, in general, \( \neg p \) is explained as \( p \rightarrow \neg q \), where \( q \) is a proposition for which refutation has been specified and for which we have a proof. So, perhaps we can apply this to the empirical case in a suitably modified form?

\[ \text{[NEG\#]}: \]
A TSI justifies the assertion at \( t \) of 'not \( s \)' just in case we can recognise that any TSI justifying the assertion at \( t \) of \( p \) can be transformed at \( t \) into a TSI capable of effective transformation into a TSI justifying the denial of \( q \).

where as before \( p \) is the translation of \( s \) into the metalanguage.

By specifying appropriate substituends for \( q \) (i.e. some Anti-Realistically sanctioned truth) for a given type of statement \( p \), we have an explanation of \( \neg \cdot \) in empirical contexts. Instead of (7), we would then have:

\[ (9) \ (\forall t)(\forall A)(\forall s)( A \cdot t) \iff E(p \rightarrow \neg q, A, t) \]

How can we recognise in our informational state TSI\(_1\) that we are in a TSI which can be 'transformed' into a specific informational state TSI\(_k\)? Let us say that TSI\(_1\) is said to be \textit{transformable} into TSI\(_k\) if a logical restructuring of all the information available in TSI\(_1\) alone would result in a TSI informationally equivalent to TSI\(_k\).

So we recognise that the assertion at \( t_o \) of 'Ming the mouse is not in the house' is justifiable if we then recognise as valid any argument deducing from the supposition that Ming is in the house at \( t_o \) the consequence that, say, Ming is invisible. But this is utterly implausible. Since it is logically (and indeed empirically) possible that the evidence at \( t_o \) which warrants the assertion of 'Ming is not in the house' is defective, any argument purporting to prove that the contrary supposition is absurd must be quite unsound. [NEG\#] therefore cannot be an adequate explanation of the meaning of 'not'.

\[ \]
Now it might be thought that negation can after all be handled quite simply on an assertibility conditional theory of meaning by the clause:

\[(10) \text{jj('not's')} \iff \neg p\]

where, as before, \(p\) translates \(s\) and '−' is the metalinguistic negation operator. But this suggestion runs into precisely the same difficulties as the parallel suggestion involving \(s\) at (1) \(\text{jj}(s) \iff p\) - namely, the LHS is decidable where the RHS need not be.

Wright who wishes to maintain (10) at Truth Conditions and Criteria gave a non-standard reading of '−' in explaining the meaning of (10) which may be of use to us in grasping the Anti-Realistic understanding of 'not'\(p\):

'Any state of information justifying the assertion that 'not' \(p\) is warrantedly assertible justifies the denial of \(p\) and conversely.'

The suggestion, to which Wright is committed here is:

\[(11) \text{jj('not's')} \iff \text{jj('it is deniable that p')}\]

which we can write as \(\text{jj('not's')} \iff \text{jj('D(p)')}\). So the proposal is that a speaker grasps the meaning of the negation of \(p\) just when he understands the situations in which \(p\) can be denied.

But whilst looking intuitively attractive, this proposal runs into a difficulty which Appiah notices: we need a separate clause for the double negation of \(p\) whenever our proof theory outlaws double negation elimination. However,

\[(12) \text{jj('not-not's')} \iff \neg \neg p\] comes out, given (11) as:

\[(13) \text{jj('not-not's')} \iff \text{jj('D(D(p))')}\]

But (13) is neither true nor warrantedly assertible as it counsels us to assert 'not'\(\text{not's}\) whenever we lack counter-evidence to \(p\). This example actually brings out the original problem with (1) more vividly. For where \(p\) is

\[\text{Wright cannot avoid commitment to (12), as Appiah argues [loc. cit. p.95], whilstever he wishes the RHS's of his biconditionals to translate the LHS's in the truth theory. For 'not''p will then be translated both by '−\(\neg\)' by (9) and '\(\neg\)' is deniable 'by (10), if '\(\neg\)' translates p and (9) is the rule for negation, (12) must be the clause for 'not-not''p.'}\]
undecidable, ‘it is deniable that p’ is decidable. So in any given TSI, either D(p) is deniable or not (because of the decidability of the predicate D(ξ)). Then if it is deniable that p is deniable, it is not deniable that p (through decidability of D(ξ) again) and this could be either because p is true or because we lack evidence either for p or for not-p. So all D(D(p)) requires is the absence of counter-evidence to p.

This should confirm our strong pre-theoretic intuition that the only acceptable clause for negation in a meaning theory is:

(14) ‘not-s’ is true iff p is not true

(with p translating s)

If this intuition is correct, then the Anti-Realist ought to forsake the project of developing a meaning theory in terms of a central notion of assertibility (i.e. Option (A)) and pursue instead option (B) - of retaining the truth-theoretic paradigm, explicating meaning in terms of a non-Realist notion of truth. Then (14) could be more perspicuously represented as:

(14*) ‘not-s’ is true iff it is not true that p

[where true1 stands for the relevant notion of Anti-Realistic truth].

At a first approximation, p will be true1 when we can recognise of p that it will still be assertible in any TSI succeeding our present one.

This new proposal simply looks like the old one renamed, but this appearance might well be due to the crudeness of our first approximation to Anti-Realistic truth. It is too strict a demand to place on an empirical statement that it only be considered assertible when we have a conclusive verification of it, since as we've already seen, most empirical statements are
not conclusively verifiable. So, an Anti-Realist account of the assertibility-
conditions of empirical statements must therefore run something like a
Realist account:

(15) The assertion at t of s by A is justified iff A has evidence at t that s is
true.

[which can be written: \((\forall t)(\forall A)(\forall s)(J(s,A,t) \iff E(T_t(p),A,t))\)].

The assertibility-conditions of a negated statement would be:

(16) \((\forall t)(\forall A)(\forall s)(J(\neg s,A,t) \iff E(T_t(\neg p),A,t))\)

So, consider our previous example. A asserts at \(t_0\) 'Ming is not in the house',
after looking very carefully through each cupboard and hiding place that he
can think of. Let us suppose that A's investigation is as thorough as any
person's could be, but that Ming has eluded discovery by stealthily creeping
from one place to another. Is A justified in asserting at \(t_0\) that Ming is not in
the house, even though evidence subsequently uncovered at \(t_1\) shows this
assertion to have been mistaken? Surely, yes. Yet, if the Anti-Realist account
of the assertoric force of negation is correct, A, in asserting that Ming is not
in the house, is claiming that he has evidence at \(t_0\) that in any TSI extending
TSI\(_{t_0}\), that sentence will still be assertible. TSI\(_{t_1}\) is a TSI that extends TSI\(_{t_0}\),
but the sentence 'Ming is not in the house' is no longer assertible; to the
contrary 'Ming was in the house at \(t_0\)' is both assertible and true. So, A's
claim at \(t_0\) about the the evidence available to him, when viewed from the
enriched informational vantage point of TSI\(_{t_1}\) is both erroneous and
unjustified. For, whilst A had evidence at \(t_0\) for not-\(p\), he did not have any
evidence at all to support his conviction that TSI\(_{t_0}\) extending TSI\(_{t_0}\) would
continue to confirm not-\(p\). It is quite unclear how, on the Anti-Realist's
account one is supposed to recognise of a certain TSI that a certain feature of
it will persist into the future. How could one have such a confidence when our evidence is constantly changing?

The Anti-Realist might reply that the assertibility rule for negations and for the other operators, should be a probabilistic rule: thus (16) should really be:

\[(16) \ (\forall t)(\forall A)(\forall p) (j('not'p', A, t) \iff P(T_t(\neg \emptyset), A, t) > c)\]

where 'P(T_t(\neg \emptyset), A, t) > c' is to be read: 'the probability for A that \(\neg \emptyset\) is true at \(t\) is greater than (some specified value) \(c\)'. Then, if in 90% of similar searches A has found the object looked for, A might assign a probability of 0.9 that \(\neg \emptyset\) is true at \(t\) and some probabilistic rule might sanction the assertion of statements of this form if the probability of their being true at \(t\) is greater than 0.8, say. Then 'not'\(p\)' would be assertible for A at \(t_0\). The assertoric force of 'not'\(p\)' would then be: 'there is a very high probability that \(\neg \emptyset\) will continue to be assertible in any TSI extending the present one'. Similar probabilistic rules could be given for the other constants.

This approach, whilst more promising than the others, depends on the tenability of an Anti-Realist notion of truth for empirical statements. The meanings of the connectives are to be given in terms of the probability that statements involving them are constructively true.

I now want to give a reason for thinking that our understanding of the connectives in empirical contexts is in fact governed by a verification-transcendent notion of truth. Suppose a fire started by one of two faulty circuits burns down a house. Let \(p\) be the statement: 'circuit A caused the

---

25 This is fine provided we understand the metalinguistic '...', but as every extant probabilistic semantics depends upon a classical M\(L\), this does not augur well for the Anti-Realist's attempts to explicate the meaning of his QL negation operator.

26 This example comes from Dorothy Edgington's 'Meaning, Bivalence and Realism' Proceedings of the Aristotelian Society, n.s. vol. 81, 1980-1981, pp.153-173. She credits it to Ian Hacking.
fire'; \( q \) be the statement: 'circuit B caused the fire'. Assume the evidence of charred wires etc., makes it impossible to tell whether it was a failure in A or B which caused the fire. It nonetheless seems eminently plausible to regard the assertion of \( p \) or \( q \) as correct.

Now the Anti-Realist insists that a disjunction is only assertible if it is possible in principle to find out which disjunct holds. But in this case it appears as if it is not possible to find this out: had a person been suitably placed with the right equipment at the time, it would indeed have been a straightforward matter to decide whether it was A or B which caused the fire; but this is no longer the case - the OVP has passed away and we are now left with only inconclusive information which cannot be 'enriched' in the way that the Anti-Realist requires.

The Anti-Realist might wish to contest this: we cannot rule out the possibility of a contrite arsonist coming forward and explaining how he tampered with the wiring in the house so setting it up that circuit A, say, caused the fire. Although I believe this objection to be misguided, I will grant it pro tempore.

So let us imagine an experiment set up to simulate the actual fire with a scale model. Two faulty circuits A and B are again built into the wiring so that it is equiprobable, let us suppose, that either caused the fire. All witnesses, including the designer of the experiment, are gathered together in the one room. The experiment is run and eventually a fire does break out and destroys the model house. All evidence is then destroyed. There is then no possibility of evidence ever coming forth to confirm whether the fault in circuit A or in circuit B was responsible for the conflagration. Yet, the Realist contends, \( p \) or \( q \) is correctly assertible.

The Anti-Realist might still wish to protest even for the controlled case that it is not impossible that evidence might be uncovered proving (or, at least
strongly confirming, since we are allowing probabilistic interpretations of the
Anti-Realist's constants) that it was A rather than B\textsuperscript{27}.

I said above that a similar reply to the original case was misguided. The
reason is this: we know that no such evidence will be forthcoming for
precisely the same reason that we know that no evidence will be
forthcoming that Hamlet wore a wig - no such evidence can arise because we
are stipulating that it won't. The stipulation ought only to be unacceptable to
an Anti-Realist if it makes unacceptable Realist presumptions, but there
seems nothing endemic to Realism in the stipulation that all evidence might
be destroyed. Moreover, the logical possibility or low empirical probability of
deciding evidence arising will not suffice on a probabilistic semantics to
assert a given statement. We require a reasonable level of likelihood in order
to assert a statement on such a semantics, and if we have a prior probability
of zero or very close to zero for evidence emerging for either disjunct, we
cannot, Anti-Realistically, assert the disjunction.

The Anti-Realist might, assuming he concedes this point, retreat to asserting
the double negation of 'p or q', rather than 'p or q'. This move strikes me as
quite ad hoc. Why should we just accept his word that \( \neg\neg(p \lor q) \) is assertible
in such situations? An Anti-Realistically plausible case can be made out for
asserting not the double negation of that disjunction, but its negation
instead. The argument runs as follows: On one plausible model for negation
(the intuitionistic one), 'not'p' is assertible just when we have good grounds
for believing that it is impossible to produce any proof of p; we have good
grounds for believing that it is impossible to produce any evidence

\textsuperscript{27}For example, a secret film of the experiment which allows a decision to be made from the shape
of the flame and the direction of burning.
supporting \( p \) (in our disjunction) and equally good grounds for believing that it is impossible to produce any evidence supporting \( q \); hence \((-p \& -q)\) is assertible and since \((-p \& -q)\) intuitionistically entails \(-(p \lor q)\), we have good grounds for asserting the latter negation.

No doubt the Anti-Realist will take this argument to show that we ought not construe negation intuitionistically in empirical contexts. Quite so. But then how are we to construe it? If we are operating with a probabilistic intuitionistic semantics, \(\neg (p \lor q)\) will mean, roughly, 'there is a very high probability that it is not the case that it is not the case either \( p \) is assertible or \( q \) is assertible or that we can arrive at a TSI warranting the assertion of one or the other'. But the problem is that we are unable to evaluate this suggestion until such times as we have a clear understanding of the connectives \(\&\) and \(\lor\). As Prior might have said: prefixing a compound sentence whose meaning is obscure with the operator 'it is highly probable that' does not make its meaning any less obscure; to the contrary.
§ 4.5: ANTI-REALIST CONCEPTIONS OF TRUTH

The practice of deductive reasoning blocks any straightforward identification of truth with the recognition of truth. For, given that we come to know truths that we had not known before by deducing them from other truths, it cannot be that in recognising the premises in an inference as true we eo ipso recognise the truth of the conclusion; otherwise inferring the conclusion from the premises would be otiose. But then there seems to be a tension between the validity and the informativeness of deductive reasoning since: 'If in an inference the conclusion is not contained in the premises, it cannot be valid; and if the conclusion is not different from the premises, it is useless. But the conclusion cannot be contained in the premises and also possess novelty. Hence inferences cannot be both valid and useful' 1.

This problem has been dubbed 'the Paradox of Inference'. Even allowing for the metaphorical talk of premises containing the conclusion, the problem seems genuine enough.

In 'The Justification of Deduction', Dummett attempts to develop an Anti-Realist account of truth which satisfies the apparently conflicting demands on deductive inferences that they be both valid and informative. I will call the requirement that these constraints be jointly satisfied the Adequacy Constraint on Truth, (ACT).

In this section, we will examine this account, together with some alternatives. In the following section, we take up the issue raised by the

Paradox of Inference and look at the general question of whether deduction can be justified.\(^2\)

Dummett's leading idea is that the constructivistically useful distinction between canonical proofs and demonstrations and *perhaps* even Prawitz's proof-theoretic method of justifying the latter in terms of the former, can be generalised to the empirical domain in the form of a distinction between a 'direct' means of verifying a sentence and an 'indirect' means and an analogous procedure for justifying the former in terms of the latter. Now I argued in §4.3 that Dummett's putative justification for the canonical proof/demonstration distinction was inadequate. But even if it should prove impossible to theoretically motivate the proof-theoretic distinction, it might be possible to make out a general distinction between a direct verification and an indirect verification of a statement.

Dummett tells us that the 'direct' means of verifying a sentence 'is that corresponding step by step with the internal structure of the sentence in accord with that model of meaning for the sentence and its constituent expressions which is being used'.\(^3\) Anyone who understands \(p\) must envisage the possibility of verifying \(p\) directly.

An 'indirect' verification of \(p\) would be one in which sentences of greater logical complexity than \(p\) itself occur in the derivation of \(p\). In the case of intuitionistic logical statements, an indirect verification of \(p\) might involve elimination as well as introduction inferences.\(^5\)

\(^2\)Throughout this section I am indebted to Peter Menzies' discussion at ch. 3 loc. cit.
\(^3\) 'The Justification of Deduction' in *T.O.E.* p.312
\(^4\) loc. cit. p.313
\(^5\) loc. cit. p.312
How is the distinction between the two modes of verification to be precisely and informatively drawn? Empirical sentences are supposed to have canonical verification conditions just as mathematical sentences have canonical proofs. Yet the model Dummett appeals to in order to clarify and substantiate his theory is the very model that was originally invoked to precisely undermine such a theory! That model is, of course, the Quinean one of language as a deductively interrelated network of sentences. Quine envisaged the web as a seamless one which, though sensitive to experience only at its periphery, transmitted peripheral disturbances at a point to every other point by means of those interconnections.

Dummett's model differs from Quine's in two essential ways. First and foremost, it is not holistic: not all sentences are deductively connected with every other sentence (although there is a restricted range of sentences - the logico-mathematical ones - that are connected with almost every other sentence). Secondly, the web is stratified. We can impose at least a quasi-ordering upon the total set of sentences according to their conceptual complexity. Think of the web as a series of concentric circles with the outermost bounded by experience (as determined by our recognitional capacities) and the innermost comprised of logico-mathematical sentences. Each 'circle' or stratum represents a language fragment which, as we drive towards the centre, conservatively extends its preceding (outer) circle or stratum.

For any sentence p, its position will be determined by the system of deductive relations between sentences in the network that contain constituents of p and the observations, if any, which lead to acceptance or rejection of an assertion of p. Given the position that p occupies in the network, whether a verification is direct or indirect depends upon the route
taken through the network - crucially, a direct verification of p will not involve any sentence of greater logical complexity than p.

An indirect verification of p on the other hand does not have to be responsive to the logical structure of p and may involve sentences of higher logical complexity than p itself. Dummett gives as an example the deduction of an observation sentence from general laws.\(^6\)

Schematically, this seems clear enough, but vacuity threatens as soon as we try to use this model to distinguish direct from indirect verifications in actual cases\(^7\). As Quine might have complained, direct verifications of statements are supposed to be those constitutive of their meanings where our only access to the meanings of such statements is through their direct verifications. What would a direct as opposed to an indirect verification of a counterfactual or past tense statement look like? The Anti-Realist will claim that there is no real difficulty here in principle even if the details are impossible to specify a priori - for a past tense statement p, a direct verification will consist in the existence of some specific trace attesting to the event in question; an indirect verification of the same past tense statement p might consist in our possessing direct evidence for the obtaining of causally sufficient antecedents of p.

The difficulty here is that in both direct and indirect cases, we only have the present specific traces, we do not have the past event now available to us with which to 'compare' the past tense statement to see whether it is true and so to the extent that inference and theorizing are involved in both styles

\(^6\) loc. cit. p.289

\(^7\) Dummett cannot just rest with an appeal to intuition in making good this distinction. He requires a principled basis for the distinction since he wishes to vindicate the validity of indirect verifications by appealing to the validity of direct verifications in a way which is somehow similar to the way intuitionists like Prawitz justify demonstrations in terms of the reduction steps of a normalisation proof.
of verification, both types of verification are *indirect*. The present accessibility or inaccessibility to confirmation of the state of affairs which makes p true is surely a far more significant determinant of whether any putative evidence for past-tensed p is to count as 'direct' or 'indirect'.

Perhaps Dummett would say that fossil records etc can only provide indirect evidence whereas my memory of an event provides a direct means of access to that past event. Am I then required to form a picture of what it would be like to remember seeing Julius Caesar in order to grasp the meaning of

---

\[3\] It is true that we do sometimes speak of certain types of evidence for past events as 'direct' – e.g. the existence of certain taped conversations, might, a prosecutor might claim, constitute 'direct' evidence of the fact that a certain judge attempted to influence the course of justice. But we can obviously set no store by such accidents of usage – for here it is clear that the prosecutor means that the tapes offer particularly strong or even conclusive evidence for his assertions about past events: he certainly does not mean that no inference is involved, just that the obvious inference is the correct one or perhaps the only remotely plausible one. (It would be a different matter, of course, if the accused was presently engaged in such a conversation in the adjoining room and the magistrate was able to verify this now for himself – in that sense, the electronic eavesdropper might be said to give us 'direct' evidence of a (contemporaneous) state of affairs, extending our aural powers in much the same way as a microscope extends our visual powers). Dummett admits that inference is involved in the verification of just about any empirical statement. But if this is so, then theoretical assumptions of a very general kind (e.g. of temporal precedence in causal interactions) and of a correspondingly high degree of logical complexity must also be involved, which is to render such verifications, on Dummett's criterion, always indirect. Dummett does not demonstrate why it should *matter* whether sentences of higher or lower logical complexity are involved in the verification of a past tense statement and, intuitively, it seems a matter of indifference whether we happen to appeal to a statement such as a counterfactual conditional logically more complex than p in the verification of p - in point of fact, it is hard to see how Dummett can avoid *always* appealing to counterfactuals to explicate the notion of verifiability in principle that he avails himself of and thus, on his own showing, only ever producing indirect verifications of p, for any atomic p.
'Julius Caesar was bald'? For presently observable statements and token-reflexive statements of various sorts uttered by a speaker at those times and in those contexts when he is in a position to evaluate their truth, the distinction between direct and indirect verifications seems both useful and easy to effect. But the problems are, firstly, that it just does not generalise in a natural way to past tense or future tense or modal or other logically complex statements; secondly, that there seems no readily intelligible answer as to whether a counterfactual statement is logically more complex than a past tensed statement and, finally that, as Quine has forcefully shown, the verification of any empirical statement p is always a theoretically complex matter, making it difficult to associate p with any one set of confirmatory conditions rather than another which could plausibly lay claim to being canonical of its meaning.

In §4.4, we saw that it was intuitionistically plausible to regard the introduction rules for the logical connectives as perspicuously representing their meanings since those rules ran parallel to the canonical Heyting clauses.

9 Were we to follow the lead of Frege pp.235ff in which the direct verification of an atomic statement Fa proceeds by demonstratively identifying in a 'recognition statement' (a sentence of the form 'This is x' [loc. cit. p. 488]) the object denoted by a as having F, we would be left with an incoherent model of how we grasp the sense of (understand what a direct verification would be like for) past tense statements like 'Julius Caesar was bald'. For, as Gareth Evans observed, 'Dummett's model would seem to require that the determination of the truth-value of this proposition proceed -- in this case as in others -- via the present identification of an object as Caesar, followed by establishing of it whether or not it satisfies the past-tense predicate 'was bald'. (The two stage procedure would yield 'This is Caesar' and 'This was bald'). Thus we are brought to contemplate the possibility of a current identification of something as Caesar, and to think of our idea of Caesar as something that would assist us in effecting this extraordinary discovery. The model forces us to make sense either of our returning to the past, or of Caesar's surviving to the present.' The Varieties of Reference, Clarendon Press, Oxford, 1982, pp.98–99. Evans adds in a footnote that whilst Dummett concedes in the course of discussing our grasp of the claim that Frege was a great philosopher that 'questions about whether some aged man really was to be identified as the author of those works would be a distraction and an irrelevance' at Frege p. 235ff, he (clearly, I think) 'implies that they would not be a distraction and an irrelevance in the direct verification of the assertion, in terms of which it is understood. [12n p.99]
for those constants. The introduction rules were then seen to be valid simply by virtue of the meanings of those constants and a canonical proof for a complex statement consisting of introduction rules alone was, similarly, 'self-justifying'. Demonstrations of intuitionistic logical statements, on the other hand, which consisted of elimination as well as introduction inferences, were justified by showing that an EP existed for transforming them into canonical proofs\(^\text{10}\). This held in any system admitting of a normalisation procedure for proofs.

Let us assume that Dummett is right in thinking that the canonical proof/demonstration distinction can (somewhat) be made out for all intuitively valid mathematical proofs and that, as in the first order logic case, for every demonstration there exists some EP for converting it into a canonical proof. Then, for intuitionistic mathematical sentences, we can define a notion of truth as follows:

A mathematical statement will be true only if we possess an intuitively valid canonical proof or demonstration for it ... \((MT)\)

Since the converse of \((MT)\) clearly holds, we arrive at a statement of the necessary and sufficient conditions which Dummett tentatively endorses at 'The Justification of Deduction' p.313:

A mathematical statement is true iff we possess an EP for obtaining a canonical proof of it \(\Rightarrow \text{(MT)}\)

Now, if \((\text{MT})\) is to be an adequate (Anti-Realist) definition of mathematical truth, it must pass the Adequacy Constraint on Truth, \((\text{ACT})\), imposed by the practice of deductive reasoning - viz. that deductions be both valid and informative. So how do mathematical proofs satisfy the dual constraint on truth in \((\text{ACT})\)? Dummett thinks that canonical mathematical proofs

\(^{\text{10}}\) Here again, I will waive my doubts about the theoretical soundness of the canonical proof/demonstration distinction for purposes of argument.
represent an 'unproblematic case' here. The difficulty is caused by mathematical demonstrations in which the truth of the conclusion has been established only indirectly.

Now we know that if a demonstration of \( p \) is to be valid at all, we must possess an EP for obtaining a CP for it. So, if \( p \) is established as true by a demonstration \( D \), it could have been established as true by a CP which ensures the validity of \( D \). The validity of demonstrations is thus secured.

Their informativeness is established as soon as we recognise how the demonstration can be transformed into a CP by means of the EP.

I defer discussion of Dummett's attempt to evade the Paradox of Inference in the manner above until the next section.

Dummett's formulation of both empirical and mathematical truth are put forward in a very tentative fashion. Still, it is clear that he wishes to make the analogy between the empirical and the mathematical cases on the one hand, and the proof-theoretic case on the other, as tight as he possibly can.

So, he writes:

'In the case of mathematical statements, the relationship [between truth and the recognition of truth] can, if we are disposed to do so, be taken to be as close as this: that a statement is to be recognised as true only if we possess an effective means in principle of establishing its truth by direct means. But, in the general case, we cannot demand a relationship as close as this: we should have, rather, to say that we possess an effective method for arriving at a direct verification of the statement, provided that we are given a sufficiently detailed set of observations.' 12

Given that the recognition of truth in the empirical case is irretrievably dependent upon spatio-temporal (and other) contexts and that verification unlike logical or mathematical proof is not a permanent property of an empirical statement, it is not at all clear that, supposing we can make good

---

11 loc. cit. p. 312
12 loc. cit. pp. 313-314
the distinction, an indirect verification of \( p \) can always be transformed into a direct verification of \( p \). Further, to impose this demand is to suppose that all indirect verifications are deductions, which in the empirical case, seems to require that science be as Popper says it is.

Dummett is certainly sensitive to these problems. But the question is whether his proposals properly address them. He is clear that a (conclusive) indirect verification establishes an empirical statement as true. Furthermore, he allows that such statements can be true even when it is no longer possible to directly establish them as true - as in the case of an observational sentence whose OVP is in the past. It is a little puzzling why the existence of an EM for arriving at a direct verification of statements \textit{in general} should depend upon the existence of some sufficiently detailed set of observations relevant to determining their truth, for there are countless types of statements whose truth could not be established, directly or otherwise, by \textit{any} set of observations. Only if the Anti-Realism in question took the form of a reductive Phenomenalism in which declarative empirical statements fell into two classes - the observational and the non-observational, with statements in the latter class

\[\text{At WJL 6:14, I am very sceptical that this concession on Dummett's part really \textit{is} consistent with Anti-Realist scruples. If it were allowed, the problem I posed for the Anti-Realist account of disjunction at 54.4 would simply vanish: we could assert that either circuit A or circuit B caused the fire even though we cannot now determine which because we can convince ourselves that if we had been at an OVP for the fire, we would have been able to determine whether it was A or B. But then why doesn't this play equally enable us to assert bivalence for every past tense statement which we now consider undecidable? Someone who doubts that 'Either there were or there were not an even number of people who consulted the oracle of Delphi in 469 B.C.' is true would then have his doubt quashed by this new Anti-Realist who points out that had we but camped outside the oracle's temple at Delphi for the whole year and counted his visitors, we would have found out which disjunct held! Is this really what an \textit{Anti-Realist} is entitled to assert? If the difference is just that a person living now physically could not have verified the previous statement about the oracle, then we should have to admit that bivalence held for the claim that Ronald Reagan sneezed twice in his sleep on the night of July 4, 1980.}

\[\text{I suspect that the reason Dummett stresses observations in this context is that he takes the making of an observation to be the empirical analogue of effecting a mathematical construction - cf WJL 6:13.}\]
reducible to some Boolean combination of statements in the former - can the
making of observations feature generally in the truth of all empirical
statements.

At any rate, as (A) above demonstrates, Dummett at The Justification of
Deduction and also at The William James Lectures toys with a direct analogy
between a notion of mathematical truth, as presented in (MT*), and a notion
of empirical truth: just as mathematical \( \phi \) is to be counted true iff we possess
either a demonstration or a canonical proof for it, so empirical \( \phi \) might be
counted true iff we possess either an indirect verification for it or a direct
verification for it. But, he decides, this is too restrictive since it rules as not
true perfectly acceptable sentences of which we can recognise that had we
but been there to observe the states of affairs which made them true as
these obtained, we would have recognised a truth. So he opts instead for the
following conception of truth for statements in general:
A statement \( \phi \) is true at \( t \) iff we either have in our possession at
\( t \) an EP for obtaining a direct verification of \( \phi \) or \( \text{were we to possess at } t \text{ a sufficiently detailed set of observations, } \text{we would (eo ipso) possess an EP for obtaining a direct verification of } \phi \) \( \ldots \) (GT)

Now, for the reasons explained above, I take the talk of observations to be
too limited for the purposes at hand. It might therefore be better to replace
it with an information-theoretic analogue as follows:
A statement \( \phi \) is true at \( t \) iff we either have in our possession at
\( t \) an EP for obtaining a direct verification of \( \phi \) or \( \text{were we to have been at } t \text{ at an OVP, } \alpha , \text{ for } \phi , \text{ we would have possessed an EP for obtaining a direct verification of } \phi \) \( \ldots \) (GT*)

Now the counterfactual in (GT*) must presumably be understood along
intuitionistic lines. So we are to assert the right hand side of the
biconditional only if we now possess evidence that if we had at \( t \) been at an
OVP $\sigma$ for $p$ we would have had an EM for verifying $p$ and it seems correct that if we did have such evidence now we would eo ipso have evidence for $p$. So the truth of the counterfactual depends upon present evidence and if there is present evidence for the truth of the counterfactual, that surely ought also to license the assertion of $p$ itself.

But now we must ask what progress the definition makes in defining truth: in particular, what is it for us to possess evidence of the requisite kind, warranting the assertion of the counterfactual and a fortiori of $p$? It seems that the only satisfactory answer the Anti-Realist can give to this question takes the form of an explanation of the types of circumstances in which we are justified in asserting that there is a warrant for asserting $p$. But this, as we saw in §4.4, turns out to be just the same as being justified in asserting $p$ and so no progress has been made.

In reply, the Anti-Realist might seek to distinguish amongst extensions of our present TSI those that correspond to (recognisable) reality as it really is from those that are merely consistent with our present TSI, claiming that this is what he intends in saying 'there is a warrant for asserting $p$'. So, applying this to the counterfactual in \((GT^*)\), we see that when we assert a counterfactual to the effect that if we had been at $\sigma$ we would have been justified in asserting $p$, we're alluding to the way things really are. Then this is to admit that when we use this type of counterfactual, the content of the counterfactual is not exhausted by its assertibility conditions.

But now the Anti-Realist faces an awkward dilemma: \textit{Either} the putative definition of truth in \((GT^*)\) makes no progress in explicating the concept of
truth - it simply amounts to saying "p" is true iff p, which is a trivial definition of truth; or the Anti-Realist must admit that there is more to the meaning of a counterfactual than can be communicated through its assertibility conditions and more to the meaning of 'warrant' than can be communicated through the assertibility conditions of 'there is a warrant'.

There are further, closely related, problems with (GT*). Let us assume for the sake of argument that truth cannot outrun our most refined means of establishing truth. Then we can still doubt that the biconditional in (GT*) really does state necessary and sufficient conditions for p to be true. For, to begin with, we can doubt whether the RHS of the biconditional really portrays a recognisable condition: how do we know when we are at an OVP for a given empirical statement or when, for observational statements, we do have a sufficiently detailed set of observations? The fact is that, as we've seen before, for just about any ordinary empirical statement, we can imagine evidence coming to light which would overturn our confident assertion of it. What seemed like an OVP would then in fact not be one. Even if we idealise in the direction of maximum perceptual attentiveness, computational and deductive carefulness etc - i.e. posit ideally rational (human) cognizers - we are still at the mercy of our evidence which can be partial and distorted, making it appear overwhelmingly certain that p at t₀, and subsequently, equally certain that not-p at t₁.

So the proponent of (GT*) faces another dilemma: either our OVPs are recognisable or they are not. If they are not, they are no better than Realist truth-conditions and to pair them with recognisable truth-conditions is demonstrably wrong. If they are, then the RHS of the biconditional in (GT*) is logically independent of the LHS. For the twin facts that the evidence for
empirical p seems optimal even to an ideally rational subject who is placed at an OVP for p and that it satisfies a certain structural constraint (viz. that it is effectively transformable into a verification of p that proceeds in accord with p's structure) do not themselves suffice to ensure that new evidence will not come to light overturning our warrant for asserting p. Unless we are to so index truth to our TSI that it becomes a property of statements that may be lost, the definitions (GT) and (GT*) seem to be materially inadequate.15

§4.5.2 Revisionism:

As we saw in §4.2, Dummett argues for a revision of classical deductive practice on the grounds that classical modes of reasoning non-conservatively extend the atomic fragment of language relative to the (non-Realist) conception of truth we actually acquire in learning a language.

Now Dummett does not think it unproblematic to inflate a proof-theoretic constraint into a general epistemic constraint. For he admits:

"it is not clear that the appeal to the concept of a conservative extension is licit at all"16

and goes on to claim that if we did have a conception of truth for the atomic sentences of our language according to which those sentences had determinate (classical) truth-conditions, it would be a matter of indifference whether the introduction of new vocabulary subject to the inference rules

15 If, as I argued in §1.2, Frege was right in holding that truth was indefinable we have strong a priori grounds for rejecting the adequacy of any definition of truth – Realist or Anti-Realist. But the criticisms I levelled at (GT*) are meant to be independent of the general Fregean point even though this last criticism – that we may not be able to recognise when we are in an OVP for p – obviously recalls Frege’s point.

16 loc. cit. p.316
conservatively extended that fragment or not. Still, he is committed to the belief that whenever we are able to derive in an enriched logic some sentence \( p \) which we could not derive in the original logic before from a set of sentences by adding to the old logic certain new sentences and inference rules, there must be something inherently unsound about the inference rules which permit this.

As we've seen, the notion of a conservative extension was used by Belnap to defend an explanation of the constants in terms of their inferential roles. TONK can only be ruled out when we are certain that we have characterised the deducibility relation completely: i.e. when we treat our characterisation of deducibility as a formal system - a set of axioms and inference rules involving \( \vdash \) - the system is assumed to express all and only universally valid sentences and rules expressible in the given notation. To demand that this extension be conservative is no more and no less than to demand that the inferential definition of PLONK be consistent with our antecedent assumptions about deducibility.

Belnap was clear that the demand for consistency in the form of the conservative extension requirement can only be justified if we do assume that we have completely characterised deducibility in our formal system - TONK can then be ruled out on the grounds that it fails the conservative extension constraint thus construed.

Belnap characterised that relation by Gentzen's structural rules. But these are independent of the specific interpretation - classical, intuitionistic, relevantist - that we place upon the connectives. So, the revisionary intuitionist cannot characterise deducibility and the notion of a conservative

17 loc. cit.
18 loc. cit. pp. 316-317
extension in Belnap's way if he is to argue that addition of the classical connectives to the atomic fragment non-conservatively extends that fragment - for, relative to a structural characterisation of deducibility, the classical constants do conservatively extend the fragment.

So the revisionist must make out a notion of deducibility in terms of the meanings of the connectives. The claim will then be, presumably, that we have a grasp of a non-Realist conception of truth in terms of which the meanings of the constants and a fortiori the properties of the deducibility relation are to be given. In this vein, Prawitz argues that relative to a conception of intuitionistic truth, classical rules such as DNE and NCD induce a non-conservative extension; although those same rules conservatively extend a fragment relative to a notion of classical truth.19

19 Cf. 'Ideas and Results in Proof Theory' and Natural Deduction: a proof theoretical study, Almqvist and Wiksell, 1965.
Now, on the assumption that the collection of intuitionistic inference rules that comprise the ND system $I$ completely explicate the essential properties of the relation of intuitionistic deducibility that holds between derivable wffs in $I$, it is no doubt correct to argue, as Prawitz does, that the classical ND system $C$ non-conservatively extends $I$. But what justifies this crucial assumption? Intuitionistic Predicate Logic has not been shown to be complete and it is an open question whether it in fact is so. Whichever this state of affairs obtains, it is open to the classicist to argue that one reason for this situation might be that $I$ lacks an acceptable account of negation, say - one which would justify DNE - and that were the intuitionist to add DNE to $I$ in order to obtain such an account, $I$ would expand into $C$. Lacking a Completeness Proof, the intuitionist has no demonstration that our Realist-inspired practices really do non-conservatively extend our assertoric use of atomic sentences.

However, even if a Completeness Proof were forthcoming for IPL, Dummett's argument for revising classical deductive practice would still not go through. For, as we saw in §4.2, it begs the question against the Realist who cites that practice as part of the evidence for our possessing a classical conception of truth.

Neil Tennant has suggested an argument for revisionism which does not appear to involve any explicit petitio against the Realist. Instead of assuming in a question-begging fashion that speakers grasp non-Realist truth-conditions, the Anti-Realist can try to lay down plausible constraints on deducibility which classical systems fail. Prominent amongst these is this:

---

Whatever notion of truth we deploy to explicate deducibility, it must yield a warrant for asserting \( \Psi \) wherever it yields a warrant for asserting each of \( K_1, K_2, \ldots, K_n \) in the derivation \( K_1, K_2, \ldots, K_n \vdash \Psi \).\(^{21}\)

The argument now runs:

(1) Valid arguments must preserve some desirable semantic property. The classicist claims that this is classical truth, the intuitionist claims that it is intuitionistic truth or provability.

(2) If we take the most primitive fragment of our language - that involving observational statements - we find that whenever we can deduce an observation statement from other observation statements, we could, had we been at an OVP for it, independently and directly have observed that the deduced statement was true. So (WAC) is met in the case of the most primitive fragment of our language.

(3) Now let us compare the classical and intuitionistic ND systems \( C \) and \( I \) and ask for each system whether it satisfies our adequacy constraint (WAC). The answer (claims Tennant) is that whilst \( I \) satisfies (WAC), \( C \) fails to satisfy it:-

There is no known classical method, \( M_c \), for transforming warrants for asserting \( K_1, K_2, \ldots, K_n \) into warrants for asserting \( \Psi \) in the derivation \( K_1, K_2, \ldots, K_n \vdash \Psi \). Letting 'v' denote 'we have a warrant for asserting...', we can represent this diagramatically as

\[
\begin{array}{c}
\square \quad \ldots \quad \square \quad M_c? \quad \square \\
K_1, K_2, \ldots, K_n \quad \rightarrow \quad \Psi
\end{array}
\]

Take a simple example to illustrate this - the derivation of \( q \) from \( p \) in \( C \) by the use of non-constructive dilemma (NCD). Suppose then that we have established \( p \vdash q \) by (NCD) and that we have a warrant for asserting \( p \):

\[
\begin{array}{c}
\square \quad \ldots \quad \square \quad \square \\
p \quad \rightarrow r \quad \neg r \quad (1) \quad \neg r \quad (1) \\
q \quad (1) \quad q \quad (1) \\
\end{array}
\]

Sub-proof (a) has \( r \) as an undischarged assumption; sub-proof (b) has \( \neg r \) as an undischarged assumption. Yet, the classicist claims that one is still justified in asserting \( q \) on the basis of a rule he takes to be sound - viz. (NCD).

---

\(^{21}\) Cf 'Anti-Realism and Choice of Logic' p.41: '... warranted assertability is the only licit property whose preservation by valid argument the anti-realist is entitled to give.'
(5) The intuitionist's criticism of the classicist's claim is this: just because there is no demonstration that \( \text{(NCD)} \) does in fact preserve warranted assertibility, we have no demonstration that the rule really is sound\(^{22}\).

(6) Thus since we have no evidence that classical logic does satisfy \( \text{(WAC)} \), we should revise it in favour of intuitionistic logic which does satisfy this constraint, since we have no guarantee that classical modes of reasoning are sound whereas we have such a guarantee for intuitionistic logic\(^{23}\).

Is this a non-question-begging argument against classical logic proceeding from premises that any rational person ought to accept? I contend that it is not. The problem centres around the interpretation of the adequacy constraint \( \text{(WAC)} \) and in particular of the term 'warrant'. We saw above that if the Anti-Realist is not to admit that his definition of truth in \( \text{(GT}^* \text{)} \) is trivial, he must agree that there is more to the meaning of 'warrant' than can be communicated through the assertibility conditions of 'there is a warrant'.

**What more?**

The classicist will want to know why in the proof of \( q \) in the example above at (4) by (NCD) we do not have a 'warrant' for asserting \( q \). Presumably Tennant's reason is that it is because we do not in general have a ground for asserting \( r \) or not-\( r \). But it seems to me that the classicist has a convincing reply at this point:

For him, to possess a 'warrant' for asserting \( p \) is to possess evidence strongly confirming \( p \), evidence making \( p \) highly probable. Now in classical probability calculi, \( p(r) + p(\neg r) = 1 \) and classical inferences preserve degree...
of probability. So we need to show that if we have a (probabilistic) warrant for asserting \( p \), we have, by using \( (NCD) \), a warrant for asserting \( q \).

Assuming the probability of \( p \) is 0.9 say, we would need to show that the probability of \( q \) is at least as high. But this is not difficult to show classically. We would need the joint probability of \( p \) and \( r \) and the joint probability of \( p \) and \( \neg r \) and we would then have to show that the sum of the probabilities of \( q \) conditional on \( p \) and \( r \) together with \( q \) conditional on \( p \) and \( \neg r \) was at least 0.9. Given such a demonstration, would we not have shown that, from the classical viewpoint at least, \( (NCD) \) precisely does preserve 'warrants' for asserting statements?

Tennant will certainly want to reject this interpretation of 'warrant' but it is not clear to me that there can be any non-tendentious grounds for doing so. One thing that we should note in this context is the convenience, for the Anti-Realist's purposes, of the term 'warrant'. In empirical contexts, the term can function as an empirical analogue to 'proof'. Then, just as one can define truth for mathematical \( p \) intuitionistically as the existence of a proof of \( p \), we can define truth for empirical \( p \) intuitionistically as the existence of a warrant for \( p \). As in the case of proof, one can assert that there is a warrant for asserting \( p \) if and only if one already has one (or has an EM for obtaining one). As we saw above, it is very difficult to obtain a definition of truth in terms of TSI's, since to judge that the definition '\( p \) is true just when in all extensions of the present TSI it is still assertible' is adequate is to ignore the fact that we are saying no more than that it is compatible with all our present and foreseeable information that \( p \) is true when, for all we know (at least for contingent \( p \)) \( p \) may not be true at all. But to say 'there is a warrant
for p' disguises this difficulty - it obscures the fact that we are appealing to how things really are independently of how we judge them to be\textsuperscript{24}.

It therefore seems to me that Tennant's argument for revisionism likewise fails.

\textsuperscript{24} To say 'There is a warrant' is not the same as saying that we actually have a warrant - the quantifier ranges over warrants really available. It just seems to be a linguistic accident that we have such a word, though a very convenient one for the revisionary intuitionist! The definitions of the connectives could run parallel to the definitions of the intuitionistic constants. In the latter case, the definitions are framed in terms of proof-conditions, in the former, more general, case, in terms of conditions for possessing 'warrants': thus, we have a warrant for asserting \( p \rightarrow q \) iff we know how to transform any existing warrant for \( p \) into a warrant for \( q \); we have a warrant for asserting \( \neg p \) iff it is not the case that we could have a warrant for asserting \( p \) (i.e. there are conclusive grounds for ruling out the possibility of a warrant for \( p \) emerging).
§4.6: CAN DEDUCTION BE JUSTIFIED?

§4.6.1: The Paradox of Inference

The Paradox of Inference arises when we try to reconcile the twin demands on deduction - that it expand our knowledge and that it be valid. Dummett expresses it thus:

"For there to have been an epistemic advance, it is essential that the recognition of the truth of the premises did not involve an explicit recognition of that of the conclusion ... For the demonstration to be cogent, on the other hand, it is necessary that the passage from step to step involve a recognition of truth at each line."\(^1\)

Now, on a Realist theory of truth, it is not clear that there is a genuine paradox or even a genuine problem in reconciling the epistemic fruitfulness of deductive inference with its validity. Still, many philosophers sympathetic to Realism have held that deduction cannot result in a real increase in knowledge. Thus, Plato at Meno 82a - 86c argued that we come to know through deduction only what we have known all along. Similarly, Mill alleged that all deductive inferences were veiled petitio principiis, since the conclusion expressed the 'same assertion' (albeit in disguised form) as that expressed by the premises - thus, the conclusion could not contain any new information over and above that contained in the premises. Mill's (and Plato's) view flies in the face of the facts - for, as Jonathon Lear has observed:

"The plea that the premises express the same "assertion" as the conclusion is futile. Even if one allows the problematic thesis that two distinct sentences may express the same "assertion", the possibility that someone who knows the premises may construct the proof to learn the conclusion refutes decisively the contention that premises and conclusion express the same "assertion"."\(^2\)

\(^1\) 'The Justification of Deduction', loc. cit. p.313

\(^2\) Aristotle and Logical Theory, loc. cit. p.86
Epistemically novel results can be obtained through deduction. On pain of denying this obvious fact, the Anti-Realist cannot identify truth with the explicit recognition of truth.

We saw in §4.5 how Dummett proposed to deal with the Paradox of Inference. On analogy with canonical proofs and demonstrations, we were to distinguish between the direct or canonical verification of a statement and an indirect verification of it - indirect verifications were effectively transformable into direct verifications under certain recognisable contingently favourable circumstances. Thus \( p \) was held to be true at \( t \) just when we either had in our possession at \( t \) an EP for obtaining a direct verification of \( p \), or we would have had, had we been suitably placed at an OVP for \( p \) at \( t \). A deductive argument with \( p \) as conclusion could be both valid and informative in the case where the canonical means of verifying \( p \) was not by means of such an argument - the argument would be valid insofar as it showed how to transform canonical warrants for asserting the premises into a canonical warrant for asserting the conclusion; it would be informative to one who, though understanding each of the premises and the conclusion, had not envisaged the possibility of an EP for transforming the aforementioned warrants in the prescribed way into a canonical warrant for asserting the conclusion.

Now, there is little difficulty in seeing how this type of case could usefully expand our knowledge - someone who understood \( p \) need never have discerned a means of so rearranging the canonical grounds for asserting the premises from which it is derived that we arrive at a canonical ground for asserting \( p \); the difficulty will be to show that this \textit{can} in fact be done - i.e.
that the deduction with p as conclusion is in fact valid. Dummett addresses this problem in 'The Justification of Deduction' ('JD') and cites Euler's solution to the Königsberg bridge puzzle as an example of an indirect verification. But, there is another type of case where it is the informativeness of the inference which seems less obvious. Dummett takes as his example here (&I) and says: 'On any possible view, it is part of the meaning of 'and' that a conjunction cannot be established save by establishing its two constituents; hence there can be no problem about the essential role of the rule of conjunction introduction in anything serving as conclusive grounds for a conjunctive statement ... The truth of a conjunction ... simply consists in the truth of the premises from which it is inferred by means of and-introduction, and so the recognition that it is true is not the recognition of a property which it had independently of the possibility of inferring it in that way.\(^3\)

I take it that what Dummett means here is that the canonical (meaning-constituting) verification of the conclusion (p & q) corresponds precisely to the deduction of that conclusion from its premises p and q: to verify (p & q) just is (i) to verify p; (ii) to verify q; (iii) to realise that one has done both (i) and (ii).

Someone who has achieved a verification of both premises thus has done all that is necessary to verify the conclusion even if in fact he does not recognise this - epistemic advance is achieved as soon as he does recognise this.

Now if we are to evaluate Dummett's proposals properly, it is crucial that we distinguish (a) canonical derivations from (b) canonical proofs. Someone who having canonically proved A then goes through the steps of a derivation of B from A and concludes that B, has not thereby constructed a canonical proof of B. A valid derivation of B from A together with a canonical proof of A does not, in other words, constitute a canonical proof of the conclusion B. We must bear this in mind in what follows.

\(^3\) loc. cit. pp.312-313
Dummett uses the Königsberg bridge example to illustrate how an indirect verification of a statement is yet valid. Dummett emphasises that one who makes an observation selectively discerns certain patterns in his sensory information (and overlooks others). Now it seems possible that Klaus could observe that Fritz had crossed every one of the Königsberg bridges without observing that he had crossed at least one twice\(^4\); had he been more observant, he could have detected this pattern directly\(^5\).

But suppose Klaus is familiar with Euler's proof. Then he knows that it is a consequence of the proof that if Fritz crossed every bridge on a certain occasion, he must have crossed at least one twice. So, given his observation that Fritz has crossed every bridge, he can deduce that Fritz must have crossed at least one twice.

Let \(p\) = the proposition that Fritz has crossed every bridge; \(q\) = Euler's theorem that any \(S\) who crosses all the (Königsberg) bridges crosses at least one twice; \(r\) = the proposition that Fritz has crossed at least one bridge twice.

The inference \(p, q \vdash r\) then comprises an indirect means of establishing \(r\). For the direct means of establishing \(r\) is to simply observe that it is so when it is - and Kurt could have observed this had he been more attentive.

Now it is comparatively trivial to show that Kurt has expanded his knowledge by feeding his observations at \(p\) into Euler's procedure at \(q\): he did not notice at \(t_0\) the time of Fritz's walk, that Fritz had crossed some

---

\(^4\) Indeed, it seems possible that this could be the content of Fritz's self-observation.

\(^5\) Cf 'in a given case we may have verified the premises without having noticed or recorded the whole route in detail' loc. cit. p.314.
bridge twice, but he knows now, having worked through the argument above, that he must have.

However, it is not trivial to show that the indirect verification deploying the inference \( p, q \vdash r \) is valid, if that requires us to show, as the Anti-Realist demands, that this verification remains faithful to the meanings conferred upon those statements by their direct verifications.

Dummett indicates how this can be done. The suggestion is that an indirect verification of \( A \) is valid just when we either have in our possession an EP for obtaining a direct verification for \( A \) or we recognise that had we but had a sufficiently detailed set of observations, we would have obtained an EP for acquiring a direct verification of \( A \). And it seems that Kurt has such an EP available to him - it is, to express it very roughly, this:
(1) Take the sequence of observations which verify the premise 'S crossed every bridge'.

(2) Feed these observations into the general procedure implicit in Euler's proof.

(3) Euler's proof will then restructure the original sequence of observations as a sequence directly verifying the conclusion.

In this manner, the indirect inference p, q |- r is to be vindicated as valid.

We can set this out more explicitly as follows: Let us represent Euler's theorem in q as \((\forall x)(\mathcal{O}(x) \supset \Psi(x))\); the proposition that Fritz crossed every bridge in p as \(\mathcal{O}(x)\); the conclusion that Fritz crossed some bridge twice in r as \(\Psi(x)\). Then the canonical derivation of r from p and q can be represented thus:

\[
\begin{array}{c}
\mathcal{O}(t) \\
\Psi(t) \quad (\forall x)(\mathcal{O}(x) \supset \Psi(x)) \quad (VI) \\
\nabla_p \quad (\forall x)(\mathcal{O}(x) \supset \Psi(x)) \quad (VI) \\
\mathcal{O}(a) \quad \mathcal{O}(a) \supset \Psi(a) \quad (VE) \\
\Psi(a) \quad (\supset E)
\end{array}
\]

We now have all the materials necessary for finding a canonical proof of the conclusion r (= \(\Psi(a)\)); the canonical derivation, to recall, is not by itself such a proof - what it does do is provide an EP for transforming canonical proofs of the premises into a canonical proof of the conclusion:

(1) We have canonical proofs for p and q:

\(\nabla_p\) - The observation that Fritz crossed every bridge.

\(\nabla_q\) - Euler's proof, a self-evident logical truth.

(2) An indirect verification consists in a canonical verification of the premises together with the derivation of the conclusion.

(3) We have a canonical derivation of r and canonical proofs of the premises p and q.
(4) So we can obtain a canonical proof of the conclusion \( r \) by transforming the canonical evidence for \( p \), in accord with the steps of the derivation, into canonical steps for \( r \) according to the RHS of the canonical derivation of \( r \).^6

Dummett's argument is ingenious and, if correct, surely does constitute a partial justification of deduction, insofar as we use deductive arguments in the context of making observations. Moreover, there is nothing in his argument which seems to require that only intuitionistic or other constructivistic inferences be used in one's derivations.\(^7\) So, a Realist should be just as enthusiastic for it to succeed as an Anti-Realist. I believe that he is in fact right. But just because the belief that deduction cannot be justified is so deeply ingrained in contemporary philosophical thinking, we need to examine the argument closely.

For a start, we can note, as Dummett himself does, that the indirect verification can only actually yield an EP for obtaining a direct verification if implemented at the time of observation. Later on, Kurt might remember that Fritz had crossed every bridge but his memory might not extend to the order in which the bridges were crossed or the point at which Fritz started - just the information that would be required as input to the reorganisation of that information effected by Euler's proof. Further, if this could be the content of Kurt's memory, it seems that it could also be the content of Kurt's contemporaneous observation of Fritz's walk - if so, the same result would

---


^7 As we saw at S4.5, Tennant's accusation that classical inferences do not preserve 'warranted assertibility' appears groundless if we take a neutral stance on what might constitute a 'warrant' for asserting a statement.
hold: the observation would not provide information detailed enough to serve as input to the Euler procedure.

As WM demonstrates, Dummett is well aware of these problems—he recognises that empirical verification is in general indexed to time and place and is, as in the example above, often both ephemeral and fortuitous. Still, if the notion of verifiability in principle is to have any application at all, it is surely applicable here. Kurt can surely recognise this feature of his own observations of Fritz’s walk: that had he himself been more attentive at t₀, he could have observed there and then what now requires an argument to establish—namely, that Fritz crossed some bridge twice.³ Furthermore, it cannot be a criticism of the soundness of Dummett’s proposal that, as above, an observation might not provide perceptual information sufficiently detailed to serve as input to the Euler procedure. For if it were to be inadequate in this way, it could not constitute a canonical warrant for p in the indirect verification above, hence Dummett’s procedure is inapplicable. The most that one could complain of is that his procedure fails to generalise beyond the observational case.

I argued in §4.5 on (vaguely) Quinean grounds that it was very implausible to think that sentences generally had a set of evidential consequences they could call their own, ones which were canonical of their meanings. Although I am prepared to acknowledge the usefulness of a direct/indirect verification distinction in the observational case together with the demonstration of the soundness of indirect verifications afforded by Dummett’s argument, I do not

³ Moreover, we can imagine the Königsberg bridges under constant surveillance by video cameras so that no one can cross all the bridges without having the fact that they crossed some bridge twice recorded on film. The OVP for directly verifying r in the inference above would thereby be indefinitely extended into the future.
see that it is plausible to extend this distinction beyond the observational case. For in regard to statements about the past or other minds or ones involving counterfactual hypotheses, indeed all non-ED statements, all our evidence is indirect. It therefore seems to me that Dummett's demonstration of the soundness of indirect verifications is quite independent of his allegiance to the Anti-Realist doctrine that the meaning of a statement is given by its direct verification conditions ... and, in point of fact, it seems to me that there is some conflict between the two theses.

Dummett acknowledges that the Anti-Realist has to show how the indirect verification of p can be in accord with the meaning of p as this is determined by p's canonical verification conditions. But to allow new grounds for asserting p which cannot be reduced to or identified with the original grounds for asserting that statement is, from the Anti-Realist's viewpoint, to modify the meaning of p in a way that is indeterminate in relation to its original meaning - it thus becomes a genuine question whether this new verification is in fact sound, given the Anti-Realist's claim that the soundness of our verification procedures derives from the meanings of the statements upon which they operate.

How is the direct/indirect distinction to be generalised beyond the observational case? We have seen that it is not at all obvious how to draw this distinction. But in the absence of a principled basis for the distinction, every type of verification of a statement becomes a meaning-constituting one. The only alternative to admitting on the strength of this that every statement is systematically ambiguous is to embrace Quine's holistic form of verificationism, as Dummett clearly recognises at WJL6.
§4.6.2 The justification of fundamental logical laws

Dummett wonders how a fundamental logical law can be criticised or justified. In JD he presents 'an apparently convincing argument' to the effect that such laws cannot really be justified at all - that they are immune from rational evaluation. The argument is, briefly, that when we reach basic laws which are not accepted because they can be derived from more basic ones, the only possible justification for them would be a semantic rather than a proof-theoretic one. So we would have to show that such laws are truth-preserving. But:

'in demonstrating soundness, we must use deductive argument and in so doing we either make use of those very forms of inference we're supposed to be justifying or we make use of the inference rules we'd already justified by reduction to our primitive inference rules ... we should therefore either eventually be involved in circularity, or have embarked upon an infinite regress.'

Dummett cites Nelson Goodman as a leading protagonist of this argument. He observes that there is then a puzzle concerning the status of soundness and completeness proofs for logical systems embodying such laws - for the soundness proof purports to show that these laws are in fact valid and the completeness proof that any valid inference may be effected by iterated applications of these laws. So the puzzle is that we have very good candidates for justifying primitive inference rules - the soundness and completeness proofs for logical systems embodying these rules - 'in the face of an apparently convincing argument that no such justification can exist'.

Dummett sets out to show that this 'apparently convincing argument' is flawed.

9 loc. cit. p.292, italics mine.
10 loc cit 291
11 loc cit 295
I shall call the thesis that Dummett opposes the thesis that fundamental logical laws enjoy indemnity from criticism - (ICT)\(^{12}\). Dummett notes that the question of justification can be raised at three levels\(^{13}\):

1. We may wonder whether a particular inference is valid and be convinced that it is by seeing that a proof with it as conclusion can be constructed by the use of simpler inference rules acknowledged as valid.
2. We may wonder whether a simple, basic form of inference is correct (or similarly whether a whole systematisation of a certain area of logic is) and here soundness and completeness proofs purport to show that the inference (or the system of inferences) is in fact correct.
3. We require an explanation not of why we should accept certain forms of argument or canons for judging forms of argument, but of how deductive argument is justifiable at all.

Now (1) is philosophically unproblematic while ever adequate answers can be provided to (2) and (3). As it stands, although (3) is obscure and Dummett admits this -

'It's not obvious what is meant by saying "deductive reasoning is justifiable". So we seek simultaneously an elucidation of that proposition and an explanatory argument showing what makes it true.'\(^{14}\)

- I take it that his solution to the Paradox of Inference gives a partial answer to how, at least for observational statements, 'deductive argument is justifiable at all'.

The notion of an 'explanatory' argument in the above quotation is a technical one in this context: 'explanatory' arguments are to be contrasted with 'suasive' arguments. In a suasive argument, the epistemic direction must coincide with the consequential one: on the assumption that S already believes the premises \(\emptyset\), we wish to persuade him that the conclusion \(B\) is true in the argument from \(\emptyset\) to \(B\). In an explanatory argument, the epistemic direction may run counter to the consequential one - we assume that S

---

\(^{12}\) Dummett assimilates (ICT) to a holistic thesis, but the connections between it and holism are quite obscure to me, especially since the most renowned holist precisely denies that any statement enjoys indemnity from criticism.

\(^{13}\) loc. cit. p.297

\(^{14}\) loc. cit. p.296
agrees with us that the conclusion \( \beta \) is true and we seek to provide an explanation as to \textit{why} it is true\(^{15}\).

\( \text{ICT} \) claims that attempts to justify fundamental logical laws either by semantic methods (e.g. soundness and completeness proofs) or proof-theoretic methods are either circular or lead to infinite regress. Dummett responds that the alleged circularity is not of the usual sort: ‘the validity of a particular form of inference is not taken as a premiss for the semantic proof of its soundness – at the worst that form of inference is employed in the course of the proof’.\(^{16}\)

Let us imagine that we are trying to justify Modus Ponens (MP). Dummett agrees that MP might be used in a proof of soundness for a system employing MP. This would be objectionable if we were trying to \textit{persuade} \( S \) that MP is valid. But, Dummett contends, we are not trying to persuade our imaginary interlocutor that primitive rules such as MP are valid – we are to suppose that he already accepts them as valid and seeks an \textit{explanation} for their validity. If \( S \) does accept such rules, then the fact that we use (MP) in our soundness proof is no longer objectionable for we already engage in deductive reasoning and are already disposed to admit that the conclusion of a valid deductive argument follows from the premises. So, if \( A \) presents \( S \) with an explanatory argument with ‘MP is justified’ or, indeed, with ‘deductive reasoning is justified’, as its conclusion, \( S \) can, acting as he is disposed to, acknowledge the explanation as \textit{cogent} – i.e. as showing that the conclusion does indeed follow from the premises.

Now I agree with Dummett against his imagined protagonist of (ICT) that a deductive argument which has as its conclusion ‘MP is justified’ or even ‘Deduction is justified’ need not necessarily be circular and I claim, more

\(^{15}\) ‘It may well be that the only reason we have for believing the premises is that they provide the best explanation of the truth of the conclusion’ loc. cit. p.296.

\(^{16}\) loc. cit. p.295
strongly than this, that the thesis that all such arguments must be circular is muddled. But I think that Dummett's dialectical strategy against his opponent is misguided. The plain fact is that there are people who deny that the practice of deductive reasoning can be justified - namely, the protagonists of (ICT); likewise, there are people who deny the general validity of particular fundamental logical laws - Carroll's Tortoise denied or at least refused to believe MP, Brouwer denied LEM, Heraclitus even denied the Law of Non-Contradiction (LNC). So it appears that there are those who need to be persuaded of the truth of the general claim that deductive reasoning is justified or of the particular claim that, say, MP is justified ... and for such ones, on Dummett's own admission, no explanatory argument is of any utility in convincing them.

However, Dummett's point might be that the supporter of (ICT), insofar as he engages in the practice of deductive reasoning at all, must take it to be justified - he is committed to the very thesis that he disavows by his act of constructing an argument for it. If this is Dummett's view, then it seems close to Aristotle's in his dispute with Heraclitus: Aristotle held that LNC was, though non-demonstrable, beyond dispute. Anyone who understood anything at all must understand LNC, Aristotle contended: it was, he thought impossible to be in error with respect to LNC. Yet Heraclitus flatly denied that LNC was true. Aristotle's response is that Heraclitus is the victim of his own confusion - he genuinely believes that he believes that LNC is false, but in fact he does not believe any such thing: no one could. So Aristotle sets out not to persuade S who does not believe LNC to change his mind (since there is no such S), but to show why it is impossible to disbelieve the law.

Here, and throughout this section, I am indebted to Jonathan Lear's lucid discussion of these issues in connection with Aristotle at *Aristotle and Logical Theory*, loc. cit., ch. 6.

Dummett might similarly believe that the supporter of (ICT) is confused. Contrary to his avowals, such a one does not believe that deduction cannot be justified: he believes that it in fact is justified, this belief being amply evidenced by the fact that he makes decisions and acts upon the basis of his reasons at all - his actions belie his avowals. As Lear points out, an explanatory argument would be useless against such an S - since he explicitly denies the claim that deduction can be justified. Yet, on Dummett's view, it cannot be suasive because this would involve using the very procedures S claims lack a justification to construct an argument with the conclusion that they possess a justification. I concur with Lear in thinking that this demonstrates the inadequacy of Dummett's suasive/explanatory dichotomy.

What Dummett ought to provide is an argument that convinces any rational person (including Heraclitus, if such he be) that deduction can be justified, not one that seeks to either persuade an S who denies this or explain to an S who already believes it how this can be so.

Worries about the adequacy of Dummett's distinctions aside, it could still be held with apparent plausibility that 'explanatory' arguments are circular: for it does not matter whether one believes the conclusion 'deduction is justified' at the start or at the end of the demonstration - the point is that that conclusion can only be arrived at by that very method which has to be shown to be justified: i.e. deduction.

Now I cannot hope to discuss the adequacy of purported justifications of deduction here. But I do want to defend, with Dummett, the proponent of

19 'Aristotle's Heraclitus does not seek an explanation: he thinks he understands the law perfectly well and he thinks it is false.' loc. cit. p. 114, 25n.
such attempts against the charge of circularity or begging the question. I am assuming that our task is to persuade S who may or may not already be predisposed to believe it, that deduction is justified in some sense; to offer him compelling grounds for accepting this conclusion. Dummett, on the other hand, thinks that the task is to explain to S who is already disposed to believe that deduction is justified, how it can be that it is: to attempt to persuade an S who doubts this or, worse, an S who explicitly denies this, by means of the doubted or disputed deductive method is somehow to beg the question against him. There is no straightforward circularity involved in a justification effected by a soundness proof, say, since (as already noted): 'The validity of a particular form of inference is not a premise for the semantic proof of its soundness; at worst, that form of inference is employed in the course of the proof'.

However, if we tried to convince an S sceptical of whether deduction was justified that certain forms of inference were truth-preserving by means of such a proof, there would be an objectionable circularity involved.

I would have thought that whether an argument is circular or not depends upon the nature of that argument and not on what its intended audience happens to believe. It might be dialectically useless against one who explicitly denied the soundness of the inferences it deploys, but this does not make it circular in any sense. Similarly, whether an argument is a sound argument or not cannot depend upon what its audience believes either - for they might to a man disbelieve the conclusion of that argument and yet acknowledge it as a good argument for that (false) conclusion.

20 Now, clearly, a circularity of this form would be fatal if our task were to convince someone, who hesitates to accept inferences of this form, that it is in order to do so. But to conceive the problem of justification in this way is to misrepresent the position that we are in. Our problem is not to persuade anyone, not even ourselves, to employ deductive arguments: it is to find a satisfactory explanation of the role of such arguments in our use of language. [loc. cit. p.296]

21 This would presumably give them an incentive to search for the premise(s) that were false in the argument.
presume that when Dummett concedes that a soundness proof is circular qua suasive device, he means that it is dialectically useless.

What is it that an S who doubts whether deduction is justified or an S who believes that it is without knowing why or an S who denies that it is need to be persuaded of if they are to believe that deductive reasoning is sound? They need to be shown how it is that the premises of a deductive argument jointly necessitate the truth of the conclusion. They need to be convinced of a general truth - that whenever the premises \( s_1, s_2, \ldots, s_{n-1} \) in the inference \( s_1, s_2, \ldots, s_{n-1} \vdash s_n \) are all true, \( s_n \) must also be true\(^{22}\). Now it is perfectly conceivable that \( S \) could be in doubt about this general relation between premises and conclusion in a valid deductive argument without its being the case that he cannot recognise valid arguments when confronted with them - indeed if this were \( S \)'s epistemic plight, it would be absurd to even contemplate trying to convince him by means of argument of anything. So we are not to make the self-defeating presupposition that \( S \) must be 'deductively blind' in constructing an argument designed to persuade him that Modus Ponens or deductive reasoning in general is justified - on the contrary, we must presume that \( S \) is rational.

The circularity charge above has it that once we use Modus Ponens, say, in an argument purporting to establish that \( MP \) is valid, we have committed some sort of petitio and Dummett seems to think that this is indeed the case when we seek to persuade an \( S \) who is not already predisposed to believe it, that deduction is justified, even though the circularity 'is not of the usual sort'.

\(^{22}\) Here I am siding with Dummett in taking the semantic characterisation of validity to be the one relevant to discussing the justification of logical laws whose validity does not seem to stem from the fact that they are derivable from more fundamental laws.
I fail to see that there is any circularity at all in a deductive argument for the justification of deductive practice, or for the soundness of MP - for so long as the argument does not either explicitly or implicitly appeal to the validity of MP, it is simply beside the point, apropos the question of circularity, that it uses an MP inference. To see this, consider the following argument which purports to justify MP:

'Suppose that \( p \) is true and that \( p \to q \) is true. By the truth table for \( \to \); if \( p \) is true and \( p \to q \) is true, then \( q \) is true also. So, \( q \) must be true too.'

... (A)

Now (A) is of the form of Modus Ponens, since it proceeds:

'Suppose \( A \) (that \( p \) is true and that \( p \to q \) is true); if \( A \) then \( B \) (if \( p \) is true and \( p \to q \) is true, then \( q \) is true); so \( B \) (\( q \) is true too) ...

But, pace Haack, what is objectionable about (A) is not that it has the form (A*) of a MP inference, but rather that it explicitly assumes in its premises that the inference rule MP is valid, and since this is precisely what has to be established, it is a straightforward petitio.

Haack thinks that the reason (A) fails to justify MP is that an exactly analogous argument could be proffered to justify 'Modus Morons' - the inference rule that licenses one to infer \( p \) from \( q \) and \( p \to q \): as for MP, so for MM, we could produce an argument which had the outward form of the inference rule to be justified. She argues that it is useless to protest that this new MM-styled argument for MM is invalid because it uses an invalid inference rule whereas (A) above does justify MP because it uses a valid inference rule, for the task at hand is to justify our belief that MP is valid and MM not.

---

I agree with Haack that (A) does not justify MP because it uses a valid inference rule - it is debarred from justifying any conclusion just because it is circular: the required conclusion occurs as a premise of the argument. But I disagree with her that we cannot know whether her MM-styled argument is valid or not before deciding on which inference rules are to count as valid. For we can know that any argument using MM must be invalid irrespective of knowing which forms of argument are valid - we can know this because we can recognise that MM is itself invalid.24

I believe that Haack’s discussion highlights the dangers of taking a purely syntactic approach to the question of justifying basic logical laws but the matter is less straightforward than my discussion has so far suggested, for it would seem that Haack has a powerful line of response to my criticisms.

Haack’s claim is that any inference rule can be justified by using that inference rule in the metalanguage. Suppose that this is true. Then if we suspect or even know that some of our inference rules are incorrect, but do not know which (through inconsistencies in a formal system that uses them for example25), we would never be able to extricate ourselves from this predicament, since for any inference rules we put forward as incorrect, we can ‘justify’ them by using them in the metalanguage in the manner indicated by Haack.

24 Just take any concrete instances e.g. p=’It is raining’; q=’I am wet’ - anyone who cannot simply see that the resulting inference can be counterexemplified is deductively blind.

25 Such a situation must be distinguished from one wherein we find out that a given inference rule is incorrect by producing a ‘strong counterexample’ to it (this term is Dummett’s in W.J.L) where by this term is meant an instance of the rule with obviously true premises and an obviously false conclusion.
Dummett in *The William James Lectures* argues persuasively that Haack's crucial assumption here is not generally correct\(^{26}\). It is only in the special case of a 'programmatic interpretation' - that style of semantic theory corresponding to a Tarskian truth-definition - that the assumption that metalanguage logic and object language logic must coincide is true. For other styles of semantic theory, it is just not true that the fact that a given law holds in the underlying logic of the metalanguage of a semantic theory entails that that law can be shown to also hold in the logic of the object language. For example, if we use Beth trees to give a semantic theory for intuitionistic logic, we can prove the soundness and completeness of that logic with respect to that semantic theory even if we assume a metalanguage obeying a classical logic. Looked at from the classical standpoint, there is no risk of those intuitionistically invalid laws of the metalanguage infecting the logic of the object language since for each intuitionistically invalid classical law we will be able to prove its invalidity in the object language by means of a 'weak counterexample' (a counterexample in which though the premises are true, the conclusion fails to be true, although without actually being false). So our ability to justify or refute logical laws by appeal to a semantic theory is not impaired by adopting in the metalanguage a logic that is stronger than that which we take to hold in the object language.\(^{27}\)

\(^{26}\) Cf *WJL* 6, 7, 8 and in particular *WJL* 7:34-45

\(^{27}\) One interesting question here is whether the converse holds: is it possible to establish the validity of all the inference rules and logical laws used in the object language where the object language uses a stronger logic than that of the metalanguage? We are strongly inclined to think that this must be impossible, but perhaps we could prove the validity of some classical law using only intuitionistically valid forms of argument if we were to proceed from the premise that every sentence must be either true or false. This may seem like mere subterfuge since we will be able to convert that particular proof into one containing LEM and the other inference rules deployed in the proof. But for this to hold quite generally, it would always have to be possible to produce a logical law corresponding to every semantic principle in the way that LEM corresponds to Bivalence and, as far as I know, this has not been demonstrated.
So it would seem, if Dummett's claims are correct, that if we were to find ourselves in the situation above of suspecting or even knowing that some of our inference rules and logical laws were unsound without knowing which, we could in principle discover which ones were to blame by surrendering the use of a 'programmatic' semantic theory in favour of some other semantic theory in which the metalanguage logic could not be assumed to coincide with the object language logic. If Dummett is right (as he certainly appears to be), then, it seems that Haack's counter-objection fails: we cannot guarantee for non-programmatic semantic theories that an inference rule will be valid in the logic of the object language just because it holds in the logic of the metalanguage.

A suasive deductive argument ought not to be judged circular, then, simply because it uses deductive inferences - for even if S denies that such an argument could possibly succeed, he will be forced to revise his view if he is confronted with an argument to that conclusion for which he cannot conceive of how the premises could be true without the conclusion being true. This is just what it is for S to recognise the argument as valid.

Thus if we were able to produce a convincing non-circular argument with 'Modus Ponens is justified' or 'deductive reasoning is justified' as its conclusion, it would be spurious to object that that argument used the very law it sought to justify and was a fortiori question-begging. It would be spurious because MP would not occur as a premise explicit or suppressed anywhere in that argument. Thus in seeking to persuade a rational S of its cogency, we are not asking him to make any judgement about the general validity of MP; we are simply asking him to decide whether this particular inference is valid. He may believe with Carroll's Tortoise that one is only
permitted to infer that $q$ from the premises $p$ and $p \rightarrow q$ if one accepts the further premise $p \rightarrow (p \rightarrow (p \rightarrow q))$ and so forth, but if we present him with an apparently cogent argument with 'MP is justified' as conclusion, he will either have to admit that his former theory about the general relation between the premises and the conclusion in a MP inference was wrong or seek to discredit the truth of one or more premises on the grounds that the conclusion is definitely false. He cannot plead that we have begged the question against him by assuming the truth of the general logical law of MP, for this is just not true. We do not accept the conclusion of the argument (that MP is justified) because we have accepted the argument and also implicitly agreed to accept MP as a rule of inference. We accept the conclusion of the argument because we’ve accepted the argument as valid. If $S$ claims that we need to include as an extra premise in the argument one expressing the validity of MP, we should respond as Achilles should have responded to the Tortoise: 'the inference is valid as it stands and does not need supplementation with a premise that purportedly licenses the inference'.

So it seems to me that there is nothing objectionable about a suasive deductive argument for the conclusion that deduction is justified. The difficulty, of course, is to produce a good one. Dummett’s resolution of the Paradox of Inference is, I believe a very important step in this direction.

\[28\text{ Cf Lear loc. cit.}\]
§5.1: Kripke’s Sceptic

§5.1.1 The Sceptical Paradox

Dummett’s Manifestation Constraint (EMC) has it that:
‘The meaning of a statement cannot be, or contain as an ingredient, anything which is not manifest in the use made of it, lying solely in the mind of the individual who apprehends that meaning.’¹

So far, we have been given no compelling reason to accept this constraint.

We need to examine, then, the grounds Dummett adduces for it. Dummett makes it clear that (EMC) is somehow a consequence of Wittgenstein’s argument against the possibility of a private language. In what follows, I shall not investigate whether this is in fact so, rather, taking Dummett’s word for this, I shall enquire whether the most forceful derivation of Wittgenstein’s conclusion that a private language is an impossibility does in fact succeed. I have in mind Saul Kripke’s powerful defence of this thesis. Kripke’s derivation has the great merit of not relying on any explicitly Anti-Realist premises; so it seems ideally suited to our purposes.

According to Kripke, Wittgenstein denied that statements attributing a grasp of rules or a possession of concepts to an individual have any truth-conditions. No facts correspond to statements of the form ‘Jones means addition by “+”’.² Wittgenstein (or at least Kripke’s Wittgenstein, henceforth, KW) has a non-cognitivist view of meaning: meaning-attributions are non-factual - they express a linguistic community’s commendation of the behaviour of individuals, dignifying that behaviour as like-minded; they do

¹ ‘The Philosophical Basis of Intuitionistic Logic’, p.216.
² I will represent the meanings of expressions and also concepts by italics.
not explain how it is that a speaker acquires 'a brute inclination to answer one way rather than another'\(^3\) after he has 'grasped' a rule. Indeed, there can be no explanation of that - or at least, no explanation which simultaneously shows how one response rather than another is _justified_\(^4\).

To make out his case for this bold thesis, KW introduces the device of a sceptic who asks me to produce some fact about my past mental states or behaviour which justifies my present confident response of '125' to the question '68 + 57?'; a calculation, we are to suppose, involving numbers larger than any I've encountered to date. There must be some such limit on the size of numbers that I've actually added up until now. Yet even though 'I myself have computed only finitely many sums in the past'\(^5\), I am confident to the extent that I grasp the concept of addition, that my 'past intentions regarding addition determine a unique answer for indefinitely many new cases in the future'.\(^6\)

Yet isn't it the rule rather than my past intentions which determines these answers? KW does not question the _arithmetical_ propriety of '125' as a response to '68 + 57?'; he does not doubt the _arithmetical_ fact that \(68 + 57 = 125\). What he doubts is that I have acted in accord with my previous linguistic intentions involving '+'. For, if in the past I had used '+' to denote not the addition function but the _quaddition_ function\(^7\), my present response of '125' would, though arithmetically correct, fail to accord with my own previous metalinguistic resolve concerning that symbol\(^8\). The sceptic's

---


\(^4\) The relevant sense of 'justified' will be explained shortly.

\(^5\) loc. cit. p.7.

\(^6\) loc. cit. p.8.

\(^7\) Quaddition is defined as follows: 'x quaus y is to be the sum of x and y for x and y both less than 57, and 5 otherwise'

\(^8\) That is, an intention of the form: 'I will always use '+' to mean quaddition'
challenge is to produce some mental or behavioural fact which establishes that I meant addition not quaddition by '⁺', and thus that I should now reply '125' and not '5' if I'm to be faithful to what I meant by '⁺' in the past. KW's claim is that all introspective and behavioural evidence is compatible with there being no such fact about what I meant, hence there can be no justification for my answering one way rather than another now.

KW's sceptic does not doubt my present use of terms: '⁺' as I presently use it is to mean addition. The doubt is just whether my present use accords with my past use. Kripke emphasises that the scepticism at issue is not of the traditional epistemological sort, wherein there may exist, at least on Realist metaphysical presuppositions which the sceptic has no need to challenge, a certain sort of fact, but, if so, one which we can have no good reason to believe in. KW's sceptic essays to demonstrate that there could not be such a fact. To this end, he allows our cognitive powers to be godlike - his claim is that even given omniscience about our past actions and intentions, no fact establishing that I meant addition rather than quaddition can be uncovered; hence there can be no fact.

Now as Kripke realises, the sceptic's case can and must be generalised to all putatively meaningful uses of terms and, if such verbal behaviour is taken as expressive of determinate thoughts, to the contents of thoughts also. For otherwise it is open to me to cite as the fact which makes it true that I meant addition by '⁺' in the past, the fact that I have mastered a general procedure for computing the sum of any two numbers whatsoever. That I

---

9 Unlike Quine, KW imposes no limitations, in particular, no behaviourist limitations, on the facts that may be cited to answer the sceptic. loc. cit. p. 14.

10 Or so we are to suppose initially. This initial supposition is necessary to allow the dialectic to start.

11 loc. cit. p. 21.
have encountered (and only ever can encounter) finitely many addition problems, is, I could then claim, irrelevant, given that I have grasped the general principle underlying each separate case. The sceptic's response to this is that since each stage in the general procedure involves following a rule, nothing justifies my confidence that I meant what I think I meant at each stage - e.g. that I meant count and not quont by 'count' etc.\(^2\)

The sceptic's challenge is, I think, a fair one. Still, some might balk at specific aspects of its formulation. For instance, why is it necessary for the disputed fact about me to justify my present response of '125'? When we account for my responding as I do, aren't we providing a causal, albeit intentional, explanation for my action? Whence the need for justification? For my meaning addition by '+', to explain why I respond as I do now, we must assume that in the past I did mean addition by '+', just as I do now. But, for KW, a response will be justified just when it conforms with what I meant by '+' in the past. So the request for justification is licit.

It is far less obvious that KW's case does not trade on the acute difficulty endemic to mathematics of providing an adequate epistemology for our understanding of mathematical statements.\(^3\) Perhaps KW's case would not have looked so convincing if he had commenced with predicates applicable to concrete rather than abstract objects, ones for which empirical evidence is relevant? More to the point, had Kripke chosen a predicate with a finite extension, the fact that we cannot think of each separate instance in the special case of addition would have failed to generalise. As it stands, KW's

\(^2\) Cf 'To say that there is a general rule in my mind that tells me how to add in the future is only to throw the problem back on to other rules that also seem to be given only in terms of finitely many cases.' loc. cit. p.22

case looks as if it could be construed as a special problem about how the infinite can be represented in finite minds rather than a wholly general problem about meaning\textsuperscript{14}.

Fortunately for Kripke, both these worries can be solved with a single change. Take as our example a predicate applicable to concrete objects such as 'square'. We can, at least for the purposes of argument, assume that there are only finitely many objects in the world of which 'square' could be predicated. Then let the deviant interpretation of 'square' be $sqwore$ where an object is $sqwore$ just in case it is for all times up until and excluding the present, square, or, now and henceforth, red and wooden. So now when I correct my young son for calling his toy train 'square', who is to say that I am right so to do?\textsuperscript{15} Note also in passing that it is irrelevant for the sceptic's aims whether the interpretation I placed on '+' or 'square' was the conventionally correct one - the scepticism attaches to what I mean by those terms and not what conventional meaning they actually have.

Wittgenstein expresses his paradox thus:

'No course of action could be determined by a rule because every course of action could be made out to accord with the rule'\textsuperscript{16}.

So what are the candidates for facts that could at any time justify one response to the dictates of a rule rather than another? If it is now true of me that, even when granted omniscience, I can produce no fact about my past mental history that ensured that I meant square and not $sqwore$ by 'square', then the same thing will be true of me at some later time. Hence the


\textsuperscript{15} This is not such an unlikely occurrence. My two year-old, Alexander, is an unabashed semantic sceptic.

\textsuperscript{16} \textit{Philosophical Investigations} $\S$201.
whole idea of meaning vanishes - neither in the past nor now did I (do I) mean anything by 'square'.

Kripke looks at several candidate facts. So far as I can see, only two types of fact look remotely plausible. The first is that an agent means *square* by 'square' just when he's disposed to apply that predicate to all and only square objects. Call this the Dispositionalist Thesis (DT). The second is that one who means square by 'square' is an irreducible, sui generis mental state. Call this the Irreducibility Thesis (IT). I will briefly sketch Kripke's reasons for rejecting (DT), since, unlike the other thesis which is apt to appear mysterious, it is immediately accessible and has considerable intuitive appeal.

§5.1.2 Kripke's argument against the Dispositionalist Thesis

In its crudest form, (DT) holds that what I mean by '+' at any time is the answer I'm disposed to give at that time. Kripke's initial response to this thesis is the one he ultimately regards as the most decisive\(^\text{17}\): irrespective of what I'm actually disposed to do, there is a unique thing I should do if I'm to remain faithful to my past intentions concerning '+' or 'square'. That is, (DT) overlooks the crucial fact that the relation between meaning and future action is *normative* not descriptive. Kripke asks:

'Can I justify my present belief that I meant addition not quaddition in terms of a hypothesis about my past dispositions (do I record and investigate the past neurophysiology of my brain?)?\(^\text{18}\).

\(^{17}\) 'Precisely the fact that our answer to the question of which function I meant is justificatory of my present response is ignored in the dispositional account and leads to all its difficulties.', loc. cit. p.37.

\(^{18}\) loc. cit. p.23.
Kripke simply rests with the intuition that dispositions do not and cannot tell me what I ought to do if I'm to remain faithful to my previous intentions concerning 'square' but one might wonder how compelling this Kripkean intuition really is. Presumably, a dispositionalist will reply that since what I mean by 'square' or '+' will be determined by their respective programs as they whirr on inside my head, it is simply not true that just any response will be 'justified'; on the contrary, only those responses which correspond to the specific outputs of the program can be 'justified', in the agreed upon sense of that term.

Kripke has other arguments against (DT) which the dispositionalist might find more difficult to combat. The most significant of these is perhaps the point that my dispositions to use 'square' or '+' are only finite. Assuming physicalism, the physical mind-brain has only a finite number of possible computational states, humans live for only a finite time, etc., so how could all the infinite number of instances of the addition function be stored in such a finite brain, ready to be accessed and manifested in my observable dispositions?

Now while I think Kripke's point, properly expressed, represents a severe difficulty for (DT), it is not a simple matter to see what the problem actually is. A dispositionalist might object to the claim that my dispositions are finite on the following grounds:

'Assuredly, my actual performances involving "square" or "+" are finite - they are finite in number and they are of finite duration - and it is also a fact that their physical basis (e.g. wave patterns of neural impulses in the brain) is likewise finite (a finite number of patterns, of finite duration etc.). But how does this show that the disposition to compute the sum of any two numbers when asked for their sum or to classify a hitherto unencountered object as square is itself finite?'

\[19\] As witnessed by his remark at p.37 above.
I think there is some force to this objection. It is not at all obvious to me how one decides whether dispositions are either finite or infinite.20

Kripke’s other substantive objection to (DT) is that since agents are prone to make errors - sometimes systematic ones - in their performances, any theory which seeks to salvage what I mean from my actual responses, has no resources to answer the sceptic who takes the error-ridden output to be a perfect instantiation of a quuss-like program. I defer further discussion of (DT) until §5.3.

§5.1.3 The Sceptical Solution

Having openly acknowledged that the notion of meaning ‘vanishes into thin air’ if KW is right, Kripke outlines KW’s Humean ‘sceptical solution’ to the sceptical paradox, which has as its consequence that the notion of a private language (in a favoured sense of ‘private’) is impossible.

Kripke points out that one who accepts the sceptical paradox is not compelled to endorse KW’s ‘sceptical solution’.21 The sceptical solution begins by acknowledging that the sceptical paradox proves that there can be no fact that corresponds to statements of the form ‘Jones means addition by “+”’. There are therefore no truth-conditions that can be assigned to such a statement. Just because there is nothing which makes it true or false that the answer to ‘68 + 57?’ is ‘125’, the solution is called ‘sceptical’. But instead of concluding from this that statements attributing a concept or addition or


21 ‘But it is important to see that his achievement in posing this problem stands on its own, independently of the value of his own solution of it and the resultant argument against private language’. loc. cit. p.60.
squareness to Jones on the strength of his verbal and non-verbal behaviour are meaningless or incoherent, the sceptical solution seeks to 'save the appearances' of our linguistic practices and simply describe what we actually do. It is supposed to emerge from this that a philosophical misconception has insinuated itself into our (virgin?) linguistic practices, giving rise to a misleading explanation of the 'grammar' of locutions such as 'the steps are determined in advance by the formula' or 'Jones means square by "square"'. Properly understood, such assertions are perfectly alright.

So, having denied that any truth-conditions can be assigned to such claims, KW replaces the single question 'what must be the case for those sentences to be true?' with the two questions:

(1) 'Under what conditions may this form of words be appropriately asserted?'

(2) 'What is the role and utility in our lives of our practice of asserting or denying that form of words under those conditions?'

Our intuitive model of rule-following, fed by a philosophical distortion of our actual linguistic practice, has it that 'in grasping a mathematical rule, I have achieved something that depends only on my own inner state, and that is immune to Cartesian doubt about the entire external world'. To straighten out this confused model so that we can see the grammar of rule-following and meaning aright, 'The picture of correspondence-to-facts must be cleared away before we can begin with the sceptical problem.'

It then transpires that:

'All that is necessary to legitimate assertions that someone means something is that there be roughly specifiable circumstances under which they are legitimately assertable, and that the game of asserting them under such conditions has a role in our lives.'

---

22 loc. cit. p. 70.
23 loc. cit. p. 73.
24 loc. cit. p. 80
25 loc. cit. p. 79.
26 loc. cit. p. 77-78.
What are the 'justification conditions' and 'utility conditions' which license the assertion of statements imputing a grasp of rules to an individual? If we confine our attention to Jones by himself, just looking at his psychological states and behaviour, then 'All we can say ... is that our ordinary practice licenses him to apply the rule in the way it strikes him.'

We cannot say something like, 'He is wrong if he does not accord with his past intentions' since:

'the whole point of the sceptical argument was that there can be no facts about him in virtue of which he accords with his intentions or not.'

With our gaze fixed on Jones alone, we cannot say that he ought to have said 'S' even if he's inclined to say '125' to '68 + 57?', or criticise his brute inclinations in any way.

Considering him in isolation, then, Jones' thinking he's obeying a rule is equivalent to his obeying it. Yet this jib with our intuitive conception that thinking one is obeying a rule precisely not sufficient for obeying it. When we widen our gaze to consider Jones' behaviour in relation to that of his linguistic community, we see that Jones' authority about whether he's following a rule or not, is not to be unconditionally accepted. On the contrary, he must get a certain number of addition problems involving small numbers, for example, correct before he can legitimately be said to be following a rule. What is it in this context for a novice to produce the 'correct' response? It is just for him to produce the same answer as his teacher would produce. A teacher can judge that Jones is applying the 'right' procedure even if he comes out with a mistaken result, because he can have evidence that Jones' sub-routines of counting and carrying numbers are the

---

27 loc. cit. p.88; cf p.89.
28 loc. cit. p.88
29 loc. cit.
30 loc. cit. p.89
same as his. Consistent quus-like behaviour might license the assertion of 'Jones means quaddition by "+"' or, if persistent and erratic enough, license the assertion that Jones is severely handicapped intellectually\(^3\). We can thus provide rough assertibility conditions for 'Jones means addition by "+"' as follows:

'Jones is entitled, subject to correction by others, provisionally to say "I mean addition by 'plus', whenever he has the feeling of confidence - "now I can go on!" - that he can give the 'correct' responses in new cases; and he is entitled, again provisionally, and subject to correction by others, to judge a new response to be 'correct' simply because it is the response he is inclined to give. These inclinations (both Jones's general inclination that he has 'got it' and his particular inclinations to give particular answers in particular addition problems) are to be regarded as primitive. They are not to be justified in terms of Jones's ability to interpret his own intentions or anything else. But Smith need not accept Jones's authority on these matters. Smith will judge Jones to mean addition by 'plus' only if he judges that Jones's answers to particular addition problems agree with those he is inclined to give ... In all this, Smith's inclinations are regarded as just as primitive as Jones's. In no way does Smith test directly whether Jones may have in his head some rule agreeing with the one in Smith's head.'\(^3\)

Too much disagreement between rule-followers would mean the practice would lose its point, Kripke claims. So what is the point of such a practice as that of dignifying a type of behaviour by the title 'following the rule of addition'? Kripke contends that its utility lies in the fact that it allows agents to coordinate their actions in such a way as to realise mutual ends and projects. Thus attributing a grasp of the concept of addition to a grocer allows me, if I believe him to be honest, to entrust the calculation of the cost of the five apples I buy to him -

'Our entire lives depend on countless such interactions, and on the 'game' of attributing to others the mastery of certain concepts or rules, thereby showing that we expect them to behave as we do.'\(^3\)

Prior to Kw's sceptical paradox, we took a conditional such as 'If Jones meant addition by "+", then, if he's asked for "68 + 57?", he will reply "125", ceteris paribus' to be true because some mental state in Jones guaranteed his performance of particular additions like '68 + 57'. But the sceptical paradox

\(^3\) loc. cit. p.88-89
\(^3\) loc. cit. p.90-91
\(^3\) ibid p.93
denies precisely this. Now we must understand such a conditional contrapositively - if Jones does not come out with '125' when asked '68 + 57?', we cannot assert he means addition by '+'\textsuperscript{34}. Rule-following, according to KW, is not a causal-explanatory concept at all, but merely an honorific label we bestow on those around us who behave and respond as we do.

\textsuperscript{34} This is too strict as it stands. We must, naturally, allow for carelessness and computational error.
$5.2$: 'SCEPTICAL SOLUTIONS' AND CONVENTIONALISM

Are there any general considerations which might lead one to doubt the tenability of KW's sceptical conclusion? Crispin Wright mentions an obvious one: given the seemingly unassailable thesis that the truth-value of any sentence depends upon the meaning of that sentence together with the state of the world, how can KW stop short of a flat rejection of the concept of truth? For, if there is non-factuality in one of the determinants of truth-value, there is surely non-factuality in the resultant truth-value. So KW's position seems to lead from the unnerving but still just credible claim that there are no facts about meaning to the genuinely 'incredible and self-defeating conclusion' that there are no facts simpliciter.

Kripke claims:
"Wittgenstein has invented a new form of scepticism. Personally, I am inclined to regard it as the most radical and original sceptical problem that philosophy has seen to date..." 3.

Kripke is surely right about this: most scepticisms ask how we can know that a certain belief is justified. But KW asks how we can know what the content of our beliefs or even our commitments is. In this sense KW's scepticism is the more radical. If the inference from no facts about meaning to no facts

---


2 Here I appropriate Kripke's own words in response to KW's conclusion, vis: 'Has not the incredible and self-defeating conclusion that all language is meaningless already been drawn?' [loc. cit. p. 71]. Wright misunderstands the dialectic here. This is Kripke speaking and not KW. If it were an attempt 'to formulate the conclusion of the sceptical argument' as Wright claims (at p. 767), then it would be opposite to note that 'if you wish to reject the suitability of a given set of concepts to figure in statements apt to be genuinely true or false, this rejection cannot take the form of a denial of statements in which those concepts figure'.

3 loc. cit. p. 60.
tout court is a licit one, we can also see that it is a scepticism with the most radical consequences. Still, perhaps the seemingly unassailable thesis can be assailed? I cannot see how this can be done, for it would require a re-interpretation of the concept of truth which entirely freed it from any conceptual ties with meaning which, as Crispin Wright observes: 'looks utterly daunting and even if possible it seems that a reconstruction of the sceptical paradox is available to rob any assignment of truth-conditions to a sentence as any possible behavioural or psychological corroboration.'

If there are no facts at all, there is no such fact as the fact that there are no facts. So, given the unassailability of the truism connecting meaning and truth, the sceptic's thesis that there are no facts as to what any individual means by his use of any linguistic expression must be false since it leads to a position which is self-refuting.

Now Kripke will no doubt seek to escape this consequence by pointing to the fact that (sic!) KW's account denies that meaning attributions have any truth-conditions - such statements can only be supplied with 'justification' conditions and utility conditions. Without a firm grasp on what Kripke takes these sorts of conditions to be, we are in no position to evaluate this suggestion.

We should first note that there is a problem with Kripke's own formulation of the thesis that meaning attributions have no truth-conditions. KW emphasises throughout the development of his sceptical paradox that there

4 In fact, I think it would be better to call it the most radical form of Anti-Realism since the thesis is a metaphysical rather than an epistemological one. Any metaphysical view which rejects any application of the notion of truth at all to the sentences of a language is surely a species of the very same genus as views which severely restrict such applications, proscribing application to the verification-transcendent portions of a language.

5 loc. cit. p. 768.
is no fact about what any individual meant by an expression E in the past nor about what he means by E now. Even granted omniscience about my past mental states and behaviour, I still could not produce the requisite fact; therefore there is no such fact. This is what makes KW's conclusion a metaphysical one. Yet on page 111, Kripke writes:

'One must bear firmly in mind that Wittgenstein has no theory of truth-conditions - necessary and sufficient conditions - for the correctness of one response rather than another to a new addition problem'.
This latter claim cannot be true if the former one is. For if, as the sceptical conclusion alleges, no truth-conditions can be assigned in principle to statements imputing one rule rather than another to an agent, then Wittgenstein does have a theory about the truth-conditions of such statements - viz, that there are not any. If, on the other hand, it is still an open question whether there are truth-conditions for such statements, then it is eo ipso an open question whether there is a fact about Jones which makes his saying '125' rather than '5' to '68 + 57?' correct. Kripke cannot have it both ways. Does he try to? This is uncertain. While I am sure that the slide is unintentional, it would be very convenient to stress the non-existence of truth-conditions when attempting to establish the metaphysical thesis that there is no fact as to what I mean by 'i' (i.e. the sceptical paradox) while playing down their non-existence in the course of elaborating the (epistemic) sceptical solution. Why would the latter move be useful? Because even though there are supposedly no facts about what an individual considered by himself means by E, there might be facts about how a linguistic community responds to his uses of those expressions, which, in concert with the individual's use, might constitute the appropriate fact. Just as there can be no facts as to whether an individual considered by himself and in isolation from his club is a member of that club (there is no 'private clubbing'), yet when we widen our gaze to take in the membership list of the club and the recognition of that individual by other members as a member,

6 Kripke does shift around on this question. Witness: "... we are not looking for necessary and sufficient conditions (truth-conditions) for following a rule, or an analysis of what such rule-following consists in". Indeed, such conditions would constitute a 'straight' solution to the sceptical problem and have been rejected loc. cit. p.87. This doesn't seem to entirely rule out the possibility of finding such necessary and sufficient conditions. As we shall see, one could 'reject' the truth-conditional model of meaning here without denying that there are truth-conditions at all.

7 Here I recall Kripke's own phrase 'considered in isolation from'. In actual fact I do not find this phrase particularly perspicuous. I take it that Kripke means that when we look at Crusoe's 'narrow' mental and behavioural states, he cannot be said to follow a rule.
there can be such a fact, so it might be with rule-following: private rule-following might be fictive for precisely the same reason as private clubbing is - namely, rule-following is a relational property that an individual bears to his linguistic community (as club-membership is a relation that he bears to his club - it takes two to club, so to speak).

It is interesting to see what Kripke actually says when he addresses this issue:

"Wittgenstein holds, with the sceptic that there is no fact as to whether I mean plus or quus. But if this is to be conceded to the sceptic, is this not the end of the matter? What can be said on behalf of our ordinary attributions of meaningful language to ourselves and to others? Has not the incredible and self-defeating conclusion that all language is meaningless already been drawn?"

In response to this, Kripke, after quoting with approval Dummett's claim that the *Philosophical Investigations* repudiates the truth-conditional view of meaning of the *Tractatus*, replies:

"Wittgenstein replaces the question, "What must be the case for this sentence to be true?" by two others: first, "Under what conditions may this form of words be appropriately asserted (or denied)?"; second, given an answer to the first question, "What is the role, and the utility, in our lives of our practice of asserting (or denying) the form of words under these conditions?"

But the fundamental question is surely whether KW is entitled to this 'replacement' after he has questioned the factuality of meaning.

Now it seems to me that one could consistently hold that there is a fact as to what Jones means by 'square' and still side with Wittgenstein in 'replacing' the question 'what must be the case for "Jones means square by 'square''' to be true?' with the dual request for justification and utility conditions for this statement. One would do this just when one felt, as Dummett does, that the slogan 'the meaning of a statement is given by its truth-conditions' is ultimately untenable and that meaning must in general be given by

---

8 loc. cit. pp. 70-71
9 'Wittgenstein's Philosophy of Mathematics' p. 185 in TOE.
10 loc. cit. p. 73
justification and utility conditions - i.e. if one were a Dummettian Anti-
Realist. That their meanings are given by such conditions would say nothing
about whether it is a fact that these statement are assertible or not or
whether they had truth-conditions at all.
Kripke might say that he did not intend the justification and utility condition
model of meaning to be generally applicable - rather, in those cases where it
is applicable, we will precisely be unable to find any truth-conditions. But is
this correct? I submit that an equally plausible explanation of why we may
be forced to turn to justification and utility conditions while holding to a
truth-conditional theory of meaning in general is that a specification of the
truth-conditions of those statements might turn out to be simply
uninformative. This could happen if the fact depicted by the truth-conditions
of such statements was an unanalysable fact. Consider our club-membership
example again. Suppose that membership is granted just on the basis of
possession of a certain indefinable personal attribute \( \mathcal{O} \). We ask for the
grounds on which Jones will be selected and we’re told that he will be
selected if he has \( \mathcal{O} \). When we ask what it is to possess \( \mathcal{O} \), we’re told that S
will be said to possess \( \mathcal{O} \) iff he’s recognised by experienced \( \mathcal{O}^{rs} \) as either
already being \( \mathcal{O}^{ish} \) or at least as having the potential to be so. If we ask what
possession of \( \mathcal{O} \) consists in, this will draw blank stares - possession of \( \mathcal{O} \) of
course! So the best that we can hope for is a sketch of the types of
circumstances in which a \( \mathcal{O} \)-judgement is agreed upon by experienced \( \mathcal{O}^{rs} \)
(the 'justification conditions') and some explanation as to why, in the context
of the interests and aims of the club, \( \mathcal{O}^{ish} \)-ness is to be highly valued (its
utility conditions). Couldn’t the same be true for the rule-follower?
Suppose we can give no analysis of what following a rule consists in, no
informative statement of the necessary and sufficient conditions for
following the rule of addition, say, beyond the bare, disquotational: "Jones is following the rule of addition" is true iff Jones is following the rule of addition. How would this show that since the best we can do is cite justification and utility conditions, there are no truth-conditions to the statement?\footnote{Kripke actually defends the converse, of course – i.e. because there are no truth conditions, we must look for justification and utility conditions. But I was attempting to refute a possible Kripkean justification, framed within an overall Realist theory of meaning, for looking at justification and utility conditions alone.} Or suppose, as in our analogy, that being a rule-follower is like being a club member - there can be no rule-following in private just as there can be no clubbing in private - how would it follow from the relational status of such an attribute that it is non-factual, any more than it would follow from the relational status of being a member of a club that it is non-factual.

The problem is that it appears as if the sceptical solution fails to answer the crucial challenge issued by the sceptical paradox: namely, having denied that there is any fact as to what Jones means by '+', why doesn't the matter end there, why isn't all talk of meaning forthwith dismissed as vacuous? It simply will not do for Kripke to reply to this by describing our actual practice involving ascriptions of the rule of addition to Jones, however much on behalf of KW he seeks to correct our mentalistic misconstrual of that practice. For the sceptic's challenge was precisely to show how, given that there could be no fact as to whether Jones meant addition or quaddition by '+', this practice could possibly be coherent.
Kripke writes:

'Now if we suppose that facts, or truth-conditions, are of the essence of meaningful assertion, it will follow from the sceptical conclusion that assertions that anyone ever means anything are meaningless. On the other hand, if we apply to these assertions the tests suggested in Philosophical Investigations, no such conclusion follows. All that is needed to legitimise assertions that someone means something is that there be roughly specificiable circumstances under which they are legitimately assertable, and that the game of asserting them under such conditions has a role in our lives. No supposition that "facts correspond" to those assertions is needed.

In the face of the sceptic's challenge, this is just argument by fiat. Kripke has not proved, and I believe cannot prove, that the supposition that facts correspond to such assertions is unnecessary. Without such a demonstration, we would be well within our rights in rejecting a 'sceptical solution' from the outset, for we could argue as follows:

The replacement of truth-conditions by justification and utility conditions as determinants of the meaning of a certain sort of statement is only legitimate in those cases where there is a fact that corresponds to the statement but one that is sui generis or irreducible, such that as a result a specification of the statement's truth-conditions is uninformative. It is not that the statement has no truth-conditions - indeed if it does not have any truth-conditions then we must conclude that it is never legitimate to assert it, since we have agreed that a statement is assertible just when one has good grounds for believing it to be true - it is rather that its truth-conditions are of little use in communicating its meaning.

Now I take it that Kripke's reply to this will be the Wittgensteinian one - 'Don't think but look!' - look at the way KW develops the sceptical solution and you will see that the assumption that facts must correspond to meaning attributions is otiose, that one can be 'entitled to assert' p even though one is not asserting thereby that one believes p to be true.

12 loc. cit. 77-78. Kripke realises the awkwardness for KW's position occasioned by the denial of the factuality of meaning for he later writes: 'the picture of correspondence to the facts must be cleared away before we can begin with the sceptical problem' Indeed it must.

13 Since it is not true!

14 Here the talk of 'facts corresponding' is just another way of saying that such assertions are objectively true.

15 Witness his claim on pp.86-87: 'Following Wittgenstein's exhortation not to think but look, we will not reason apriori about the role such statements ought to play; rather we will find out what circumstances actually license such assertions and what role this license actually plays.'
So is KW right? Can we just make do with justification and utility conditions alone? For his part, Kripke has no difficulty at all in seeing the utility of such a 'non-factually based' practice:

"We say of someone else that he follows a certain rule when his responses agree with our own and deny it when they do not; but what is the utility of this practice? The utility is evident."

Kripke illustrates this 'utility' by considering the interactions between a customer and a grocer, showing how agreement in response is crucial for the mutual achievement of ends. I confess that I find it harder than Kripke evidently does to see the utility of the practice in question. There is no difficulty, of course, in appreciating the general significance or 'utility' of broad agreement in response throughout a community, at least as this touches their fundamental and everyday transactions with one another. That platitude is not in dispute. What is not at all obvious, to me at any rate, is how the mutual issuing of sequences of vocables at the point at which our actions co-ordinate can itself assist further transactions between the grocer and me if that sequence (or constituent vocables within it) does not (or do not) correspond to anything involved in such transactions (e.g. neither my 'five apples' nor the grocer's 'five apples' refers to the five apples I bought from him).

Kripke's example shows us the utility of the grocer's meaning the same as we do, but what he has to show is the utility of having that form of words in the language - it is only because he tacitly assumes the correct interpretation of

---

16 loc. cit. p. 92
17 loc. cit. pp. 92–93
18 No doubt I would have found it much easier to grasp had Kripke demonstrated how our brute inclinations succeed in landing men on the moon! The sceptic's position becomes very implausible when we consider the applicability of mathematics to the world - why should the world co-operate with our brute inclinations?
the terms 'five', 'apple', etc, that Kripke can think that he has explained the utility of the linguistic practices in question.

What of the justification conditions for attributions of rule-following, does the absence of any fact create problems for these also? We judge that Jones means square by 'square' when he applies the term as we would apply it to (what we would take to be) all and only square objects, allowing for borderline cases of course. But, according to KW, it is not because Jones and I have identical, determinate intentions to so use 'square' as to mean square that we call square objects 'square'; it is rather, contrapositively, that if we didn't call square objects 'square', we couldn't be said to mean square by 'square'. Just because Jones' use of 'square' and my use of 'square' must equally be subject to correction by others, if we are to be ascribed the concept square, there can be no private language: i.e. no justification and utility conditions for 'Jones means square by "square"' when Jones is considered 'in isolation' (i.e. if with God we could examine the contents of Jones' mental states or behaviour).

But consider our club example again. Experienced members admit Jones into their club if they consider him to be sufficiently Øisth (or to have sufficient Ø potential). We might suspect that there is no genuine property Ø which all members of the Ø-club really possessed. Still, just as there is a fact about the criterion for Ø-membership - that one be judged Øisth by other Ørs - so there might be a fact about the criterion for rule-following membership - to wit, that one be judged to be following a rule by other rule-followers. So it seems that, Kripke's assertions to the contrary, we do have a fact that corresponds to Smith's assertion of 'Jones means square by "square"' - the fact that
Smith has judged that Jones responds as he does to situations in which 'square' is applicable.

Now Smith's assertion cannot be accepted unconditionally, it is subject to the correction of other rule-followers. Granted. So the fact of Smith's judging as he does is not sufficient to make his assertion about Jones true. Then the requisite contrast between assertibility and truth arises as follows:

Smith is entitled to assert 'Jones means square by "square"' just when he judges that Jones' responses coincide with his own with respect to 'square'. This is an assertibility condition that holds good for each speaker. But, if we are to faithfully describe our actual practice of concept attribution, in accord with KW's invocation to look, not think, we need to distinguish thinking one is obeying a rule from really obeying it. The distinction arises only in the social context. Although each speaker has assertibility conditions for uttering 'Jones means square by "square"', it is only if their individual responses with 'square' coincide that this practice can have a utility for us. So we consider the community of speakers, \( C_0 \), and say that 'Jones means square by "square"' is true iff the dispositions of the majority of \( C_0 \)-members to use "square" coincide with Jones' dispositions to use "square". So whilst there may be no unrelativised fact, there is a community-relativised fact.

\[ \text{Unless of course one wants to insist against KW that Protagorean subjectivism is the reality and the withdrawing and correcting of assertions in the social context the misleading appearance.} \]

\[ \text{Naturally, } C_1 \text{-members might judge Jones to mean square by 'square' (with compensating adjustments for all the other expressions of his language). But, to transpose a lovely remark of Michael Dummett's, if } C_1 \text{'s react to our evaluations of Jones' use of 'square' with 'Here we have a hitherto unknown kind of madness', we can go away content in the knowledge that we have our own criteria of insanity too!} \]
Now Kripke considers this theory but dismisses it at page 111:
"Wittgenstein's theory should not be confused with a theory that, for any \( m \) and \( n \), the value of the function we mean by 'plus' is (by definition) the value that (nearly all) the linguistic community would give as the answer. Such a theory would be a theory of the truth-conditions of such assertions as "By 'plus' we mean such-and-such a function" ... The theory would be a social, or community-wide, version of the dispositional theory, and would be open to at least some of the criticisms as the original form."

But it just does not follow from the fact that each individual's dispositions fail to determine what he or she means that a statistical or qualitative summation of the collective dispositions of all or most individuals in a given community fail to fix what they mean. To argue otherwise is to commit a fallacy of composition.

Social convention might determine in a community what is to count as the correct answer to '68 + 57?' in exactly the same way as social convention determines within a community the fashionable way to dress. Moreover, if Jones' responses match those of his fellows over a wide range of actual and counterfactual cases involving '+' or 'square', then, in our community \( C_0 \), he means addition by '+' and square by 'square'.

Assertibility conditions and utility conditions are, as Kripke rightly perceives, closely related - if one can give no account of the point of asserting \( p \), one can give no account of the assertoric force of \('-p'\), hence no account of what we might be doing in asserting \( p \). We restored coherence with a single move which reinstates at once assertibility and utility: we provided the statement with the truth-conditions suggested by (but explicitly disavowed by Kripke) in the sceptical solution itself: 'Jones means

\[21\] It is in the context of this remark that Kripke claims that KW has no theory of truth-conditions (p111).
\[22\] In \( C_1 \) on the other hand, he might mean quadratic by '+' and square by 'square'.
addition by "+" is true iff Jones' responses agree with his fellows' with respect to '+'.

Kripke is unpersuaded by these truth-conditions. He writes:

'If Wittgenstein had been attempting to give necessary and sufficient conditions to show that '125', not '5', is the 'right' response to '68 + 57?', he might be charged with circularity. For he might be taken to say that my response is correct if and only if it agrees with that of others'. But even if the sceptic and I both accept this criterion in advance, might not the sceptic maintain that just as I was wrong about what '+' meant in the past, so I was wrong about "agree"? Indeed to attempt to reduce the rule for addition to another rule - "Respond to an addition problem exactly as others do!" - fails foul of Wittgenstein's strictures on a 'rule for interpreting a rule' just as much as any other attempted reduction. Such a rule, as Wittgenstein would emphasise, also describes what I do wrongly: I do not consult others when I add. (We wouldn't manage very well, if everyone had to follow a rule of the proposed form - no one would respond without waiting for everyone else).'

Kripke's objection seems to me to entirely fail to engage with the conventionalist's proposal - the conventionalist is not suggesting that we need to agree about the meaning of 'agree' before we can be said to agree, nor is this any objection to his proposal. The requisite agreement is, in Wittgenstein's words, 'not agreement in opinions but in forms of life' - that is, an agreement in our responses. The conventionalist is offering us a 'majoritarian' view of truth - 'p is true iff most people would agree with p'. Collusion, either implicit or explicit, is neither necessary nor sufficient for

---

23 Kripke contends at p. 112 that what follows from the justification and utility conditions for 'Jones means addition by "+"' that KW gives:

'is not that the answer everyone gives to an addition problem is, by definition, the correct one, but rather the platitude that, if everyone agrees upon a certain answer, then no one will feel justified in calling the answer wrong.'

I agree that the conventionalist position I have been opposing to the sceptical solution does not follow directly from the justification and utility conditions alone - it requires independent argument. But I have produced such an argument - namely that it is incoherent to suppose Smith's mere utterance of the sequence of vocables 'Jones means addition by "+"' to which nothing in either Smith's or Jones' behaviour corresponded, could enjoy a utility for Smith, Jones or anyone else; and it was similarly incoherent to suppose that this selfsame sequence of vocables could be asserstible in any situation except one wherein Smith had good grounds for believing it to be true; so were the statement to have no truth-conditions, it ought never to be asserted!!

24 loc. cit. 87n p. 146

25 *Philosophical Investigations* S241. Cf S242: 'If language is to be a means of communication, there must be agreement not only in definitions, but also (queer as this may sound) in judgement.'
this type of agreement: agreement in response - it is not necessary, because we could be independently inclined to give the same answers without agreeing or even disagreeing that our responses constituted 'agreement'; it is not sufficient, because we could agree on a formulation of what constituted 'agreement' and find ourselves brutally disposed to react in totally different ways. Moreover, and this is the decisive point, Kripke's own preferred solution itself depends upon this type of agreement. As he says: ‘... if there was no general agreement in the community responses, the game of attributing concepts to individuals - as we have described it - could not exist. In fact of course there is considerable agreement, and deviant quus-like behaviour occurs rarely ... almost all of us, after sufficient training, respond with roughly the same procedures to concrete addition problems. We respond unhesitatingly to such problems as "68 + 57" regarding our procedure as the only comprehensible one ... and we agree in the unhesitating responses we make.’

Kripke thus cannot raise a sceptical problem for the conventionalist's sense of 'agreement' without raising the very same problem for his own 'solution'. So Kripke's sceptical solution can be saved from incoherence only by positing communal truth-conditions for attributions of rule-following. Seen in this light, the sceptical paradox does not establish the metaphysical conclusion that there is no fact as to whether Jones means square or sqwore by 'square', but instead corrects the distorted metaphysical picture of its being a fact about Jones considered 'in isolation from' his community which makes it true that he means one thing rather than another by '+' or 'square'. The relevant fact is a social fact, not a solipsistic one - the sceptic enables us to see what the real fact is, not that there isn't any fact.

This type of agreement could be decided, as it were, by secret ballot - the votes are cast in private and whether p is true or not depends upon how many votes it polls. Thus the conventionalist will claim that the sceptical paradox reveals that we were wrong to suppose that we could recover what Jones meant by '+' through an examination of the content of his narrow mental and behavioural states: the fact is that we looked in the wrong place - inside his head - where we should have not just looked at Jones' individual narrow dispositions but at his wide dispositions in concert with those of his fellows.

26 The truth of p could be decided, as it were, by secret ballot - the votes are cast in private and whether p is true or not depends upon how many votes it polls. 27 loc. cit. p.96. 28 Thus the conventionalist will claim that the sceptical paradox reveals that we were wrong to suppose that we could recover what Jones meant by '+' through an examination of the content of his narrow mental and behavioural states: the fact is that we looked in the wrong place - inside his head - where we should have not just looked at Jones' individual narrow dispositions but at his wide dispositions in concert with those of his fellows.
But now there emerges a problem for Kripke which Wright has noted. Consider the assertibility conditions that Smith associates with 'Jones means square by "square"' once more. Kripke, as we've seen, cannot coherently deny that it is true that Smith, to the extent that he asserts this statement correctly, associates with this statement the condition that Jones' responses involving 'square' agree with his own. But what does this truth consist in? That is, what is it for Smith to make the judgement that Jones' responses agree with his own? After all, he only entertained that thought on a finite number of occasions and any dispositions he had to make that judgement were likewise essentially finite. But, it seems that only finite mental states and finite dispositions could be available to Smith to determine that he meant agree and not schmagree by 'agree' and these, if KW's argument against their status as meaning-determinants succeeds in the first place, all fail to determine that Smith meant agree and not schmagree by 'agree'. Thus there is no fact as to the content of Smith's judgement. So it seems that we are forced to deny that it is true that Smith associates with the statement 'Jones means square by "square"' the assertibility conditions that Kripke says he does - i.e. it is not true that what corresponds to Smith's claim that Jones means square by 'square' is Smith's judgement that Jones' responses agree with his own. But then there can be no truths about the assertibility conditions any of us associate with such a statement, hence there can be no assertibility conditions communally associated with the statement.

---

29 loc. cit. p.770. Wright actually tries to raise this problem without giving any demonstration that Kripke, his own avowals to the contrary, is committed to the existence of facts. But, as will become clear, if Kripke is right about the non-existence of facts, he can easily escape the problem.

30 I suspect that it is an awareness of the problem posed by this result which lies behind Kripke's insistence that there are no facts that correspond to the statements in question.
Now KW's argument against the possibility of a private language is supposed to follow from a consideration of the conditions under which we do in fact assert statements of the form 'Jones is following the rule of addition'. The argument is in brief this:

(1) There must be a distinction on any viable account of rule-following between thinking one is obeying a rule and actually obeying it.

(2) Now if we simply look at Jones' narrow mental and behavioural states, the situation is indeed one in which this distinction cannot be sustained. Jones is entitled to assert 'I mean square by "square"' whenever he simply feels he knows how to apply 'square'.

(3) But Jones' linguistic community, C₀, is in a very different position: members of C₀ will claim that Jones means square by 'square' just when they judge Jones' responses to agree with their own. Jones' putative rule-following is now subject to correction by others. If no agreement in responses could be secured, the whole practice would break down.

(4) Since we require the distinction between seeming to obey and actually obeying and since the distinction can only be made out when we view Jones' responses in relation to those of other C₀-members, there can be no 'private' rule-following or language. We can now see that the argument fails. It fails because the crucial premise (3) is false since, as we saw above, there are no facts about the assertibility conditions C₀-members associate with statements of the form 'Jones means square by "square"'. Therefore, in Wright's words:

'the premises requisite for Kripke's version of the private language argument about the community-oriented character of assertibility conditions for sentences about meaning and understanding are not genuinely factual and the same must presumably said of the conclusion, that the concepts of meaning and understanding have no genuine application to the private linguist'.

So Kripke has produced no convincing argument for the impossibility of a private language. Therefore this means of vindicating Dummett's Manifestation Constraint fails, and it seems likely that all other arguments for (EMC) require the very Anti-Realist premises a Realist rejects.

---

31 The relevant sense of 'private' is perhaps best conveyed in this passage: 'the notion of a person following a given rule is to be analysed simply in terms of facts about the rule-follower and the rule-follower alone, without reference to his membership in a wider community' loc. cit. p. 109.

32 loc. cit. p. 770.
Now the points at which I have attacked KW's sceptical solution are precisely the points which Kripke thinks are immune from such attack. So, I had better be right about this.

Kripke thinks that the assertibility conditions Smith associates with 'Jones means square by "square"' are not themselves susceptible to the sceptic's strategy. If he were correct in holding that these justification conditions need no facts to correspond to them, he might be right. But I argued that since it was pointless for Smith to sound off with the string of vocables 'Jones means square by "square"' if nothing in Jones' or Smith's behaviour corresponded to it, and since one is only warranted in asserting such a statement when one believes the statement to be true, on pain of incoherence, justification and utility conditions both demand that there be a fact that corresponds to Smith's assertion 'Jones means square by "square"'. And, I argued, there was such a fact even on Kripke's own showing - the fact that Smith judged Jones' linguistic responses involving 'square' to coincide with his own. But then it seemed that, contrary to our original supposition, it was not true that to Smith's assertion 'Jones means square by "square"' there corresponded Smith's judgement that Jones' responses involving 'square' agree with his own; and from this result it was a short step to prove the failure of the private language argument.

So, on my reading, it is the very conditions Kripke believes immune to attack - Smith's assertibility conditions for 'Jones means square by "square"' - which I think are ultimately open to the paradox. Conversely, it is the very

---

33 Kripke might wish to insist that a statement is assertible not only when one believes it to be true but also when one is entitled to believe the proposition it expresses. But we saw in §4.5 that the whole notion of what it means to be warranted in asserting a statement is opaque, at least on broadly Anti-Realist premises. Perhaps Kripke can make out a notion of 'warranted assertibility' that is more substantial than the intuitionist's, but there is nothing in his book to indicate how.
conditions Kripke thinks open to the sceptical argument which I contended were immune - i.e. the truth-conditions, or, more precisely, the alleged truth-conditions\(^{34}\) of "Jones means square by "square"." I claimed that these conditions were immune because they did not involve judgements about agreement at all. They only involved the one type of agreement Kripke cannot afford to disavow - agreement in responses.

To recall, our conventionalist truth-conditions for "Jones means square by "square"" were, roughly, ""Jones means square by 'square'" is true iff Jones' responses involving "square" match those of (most of) the other members of his community". No appeal to shared judgements of agreement is necessary to formulate those truth-conditions. But an appeal to constancy in judgements of agreement is necessary to formulate the assertibility conditions. This ought not to be surprising - it would obviously be incorrect to state Smith's assertibility conditions for "Jones means square by "square"" as ""Jones means square by 'square'" is assertible for Smith iff Jones' responses involving "square" match Smith's own since Smith might be quite unaware that Jones' responses match his\(^{35}\). So if the sceptical paradox can be turned against any factual condition on "Jones means square by "square"", it can surely be turned against KW's assertibility conditions and not the conventionalist's truth-conditions.

Now we can return to the paradoxical status of the sceptical paradox itself.

We saw before that if we were granted the truisim that the truth of a statement jointly depends upon the meaning of the statement and the state of the world, acceptance of the sceptic's conclusion that there are no facts

---

\(^{34}\) Since Kripke denies there are any!

\(^{35}\) A similar point about assertibility conditions was made at §4.4.
about what any speaker meant by any statement, a fortiori no facts about what the statement means, logically entails acceptance of the radically Anti-Realist conclusion that there are no facts *tout court*, which position is, as we saw, self-refuting. The only feasible ways to block this result were to deny the truism connecting meaning and truth, which seemed an untenable course, or to deny either that the truism or the sceptic’s conclusion or perhaps both had truth-conditions — only justification and utility conditions. But it should now be clear that this latter alternative is unworkable also. It is a necessary condition for such statements to possess justification and utility conditions that they also possess truth-conditions. We thus have good grounds for believing the sceptic’s arguments to be flawed since the sceptic’s conclusion is self-refuting. The more difficult task is to try to discover precisely where the sceptic errs.
§5.3: DISPOSITIONS, INTENTIONS AND MEANING

§5.3.1 Constraints on an acceptable answer to Kripke.

On p. 11, Kripke stipulates two conditions which any candidate for a 'straight solution' to the sceptical paradox must satisfy:

(C1) It must produce some fact that shows that I meant addition and not quaddition by '+' in the past.

(C2) It must show how that fact justifies my present response.

Now in §5.1, we discussed the relevant sense of 'justification' in (C2) and found it to be acceptable: '125' will now be justified as a response to '68 + 57?' iff my present use of '+' is faithful to my past intentions governing its use. But Kripke builds more into these two conditions than is entailed by their statements in (C1) and (C2) alone:

(C1*) The fact is to be a fact about me - i.e. about my narrow mental states or behaviour.

(C2*) The fact is to justify my present response by giving me directions determining what I ought to do in any circumstance.

(C2*) needs some clarification. Kripke writes:

'Ordinarily, I suppose that, in computing '68 + 57' as I do, I do not simply make an unjustified leap in the dark, I follow directions that I previously gave myself that uniquely determine that in this new instance I should say "125".'

1 That is, one that seeks to demonstrate how the sceptic's challenge can be met.

2 Cf: 'An answer to the sceptic ... must give an account of what fact it is (about my mental state) that constitutes my meaning plus, not quus' loc. cit. p. 11.

3 loc. cit. p. 10.
So Kripke requires that the fact justify me by somehow encapsulating ‘directions’:

'The "directions" ... that determine what I should do in each instance, must somehow be "contained" in any candidate for the fact as to what I meant'\(^4\)

Now, one might try to answer the sceptic's challenge as adumbrated in \((C_1)\) and \((C_2)\) whilst rejecting either or both \((C_1^*)\) and \((C_2^*)\) as representing unjustified construals of those conditions. But it seems to me that \((C_1^*)\) at least ought to be accepted as the only interpretation of \((C_1)\) that can meet the sceptic's challenge. To see this, consider putative straight solutions which reject \((C_1^*)\).

I argued in §5.2 that Kripke's own sceptical solution, which of course does not attempt to meet \((C_1)\) and \((C_2)\), was, as it stood, dubiously coherent, and could only be made coherent by being restructured in the form of a conventionalism which denied the distinctive claim of the original 'solution': namely, that no truth-conditions could be assigned to statements of the form 'Jones means square by "square"'. Whereas straight solutions require a fact (accept \((C_1)\)) and a demonstration that the fact justifies me (accept \((C_2)\)), sceptical solutions deny there is a fact (reject \((C_1)\)) and deny that our ordinary practice stands in need of the type of justification originally sought (reject \((C_2)\)). The difficulty was that this latter denial seemed itself to be quite unwarranted - at every point where the sceptic demanded a proof that the practice was coherent, KW simply ignored the challenge and described the disputed practice, thus presupposing that it was coherent. Moreover, the only way of proving that our linguistic practices were coherent was to supply the very facts KW claimed that we didn't need. But what were these supposed facts? They were not facts about Jones considered by himself, they were composite facts about Jones-in-

\(^4\) loc. cit. p.11.
So a conventionalist such as Crispin Wright responds to Kripke’s challenge by accepting \( (C_1) \) but rejecting \( (C_1^*) \) and accepting \( (C_2) \) but rejecting \( (C_2^*) \). The fact sought by the speaker is not a fact about me; my justification for saying ’125’ not ’5’ does not inhere in my following any ‘directions’ I gave myself but in the fact that my responses match those of my fellow speakers.

Can such a thesis be seen as an answer to the sceptic? It is an ‘answer’ in letter only, for in spirit it concedes the sceptic’s main point. The sceptic wants to know what justifies Jones’ particular application of a rule, irrespective of whether the rule Jones follows is the conventionally acknowledged one. The conventionalist responds by claiming that the question of justification can only sensibly be raised in the context of his community’s perception of meaning and rules. But this is to acquiesce in the sceptic’s negative conclusion. For the sceptic sought to attack the intuitive conception that what justifies my response ’125’ now to ’68 + 57?’ was the fact that it accords with what I meant by ‘+’ in the past ... and the conventionalist agrees with the sceptic in rejecting this model of justification.

So a conventionalist position such as Wright’s is an essentially sceptical solution in the sense that it accepts that the sceptic’s challenge is unanswerable. \( (C_1^*) \) thus seems to me to be non-negotiable for the non-sceptic who seeks to provide a straight solution. And surely, pre-Kripke at least, it does make sense to talk of my present applications of ‘square’

\[\text{footnote}{5}{it differs from KW’S official version in providing a fact, but it is not a fact about me, as is required by the very nature of the paradox.}\]
according or failing to accord with my past metalinguistic intentions regarding 'square'. For, pre-Kripke, I could have been using 'square' with a meaning that was out of step with its conventionally accepted meaning - I might have thought it applied to any four-sided rectilinear figure - and so now when I call a trapezium 'square', the question arises whether this application accords with what I understood by 'square' in the past, and is in that sense 'justified'. But on Wright's view, there is no such thing as acting in accordance with the dictates of one's own understanding, indeed this is the very model of rule-following which Wright believes Wittgenstein to have effectively demolished.

Hence we require a fact about me to answer the particular problem the sceptic poses. But do we also require the putative fact to justify my actions by encapsulating 'directions' telling me how to act in any circumstance? I take it that Kripke need not be taken to mean by 'directions' conscious verifiable instructions, just that whatever mental state or event we postulate as comprising our fact, it must have the effect of an internal imperative in relation to my actions. For otherwise: 'nothing justifies a brute inclination to answer one way rather than another.'

---

7 Although this is too weak. For electrical stimulations emanating from a chess-playing computer to the nerves controlling the movements of my arm and hand might have the same effect as certain internal imperatives I might give myself: 'move KQ4!' etc. Such stimulations surely do not pass as Kripkean 'directions'. Kripke clearly intends that we act as we do because we recognise and understand the 'directions' we give ourselves however precisely we are to construe those 'directions'.

8 loc. cit. p. 15.
§5.3.2 The Dispositionalist Thesis

The Dispositionalist Thesis (DT) claims that the fact which settles what Jones meant by ‘+’ in the past is a dispositional fact about how he would have responded had he been asked, say, ‘68 + 57?’, then. It is important to note that the allegedly meaning-constituting dispositions of Jones’ past meanings are his past dispositions. The dispositionalist will wish to distinguish:

(1) What Jones meant at time t by ‘+’ or ‘square’.

(2) Whether what Jones means at a later time by those terms conforms with his earlier meanings.

The dispositionalist claims that questions of justification and normativeness only arise in relation to (2) - i.e. whether what Jones presently means by E conforms to what he meant by E in the past. We answer this question, he maintains, by examining Jones’ past and present dispositions involving ‘square’ and seeing whether they coincide.

Now Kripke seems not to take much notice of the distinction between past and present dispositions for in summing up his objections to (DT) he writes: ‘Suppose I do mean addition by “+”. What is the relation of this supposition to the question of how I will respond to the problem “68 + 57”? The dispositionalist gives a descriptive account of this relation: if “+” meant addition, then I will answer “125”. But this is not the proper account of the relation, which is normative, not descriptive. The point is not that, if I meant addition by “+”, I will answer “125”, but that, if I intend to accord with my past meaning of “+”, I should answer “125”.

Now while I think there is a valid point being made here, Kripke’s account of the dispositionalist’s view is tendentious. The dispositionalist does not hold that if at the present time t₀ I mean addition by ‘+’, then I will answer ‘125’ at a future time t₁ if asked ‘68 + 57?’ then - he need not be committed to anything by way of prediction at all - all he is committed to is the

9 loc. cit. p.37
counterfactual claim that *had I* been asked '68 + 57?' at t₀ *I would have* replied '125' then. On the other hand, he can agree with Kripke that 'if I intend to accord with my past meaning of '+' I *should* answer "125"'.

Hence Kripke's claim:

'Precisely the fact that our answer to the question of which function I meant is *justificatory* of my present response is ignored in the dispositional account and leads to all its difficulties.'

is correspondingly tendentious. (DT) does not ignore the question of justification - it gives a simple (arguably too simple) answer to when a response is justified - viz. Jones' present response of '125' to '68 + 57?' will be *justified* just when it accords with what he meant by '+' in the past, as this was determined by his past dispositions. Prima facie then, the dispositionalist can and does have an account of justification and Kripke's consistent charges that he ignores the question of justification are baseless. Of course, taking cognizance of the question is one thing, providing a convincing answer is another and one might feel that just because dispositions are inadequate to the task of supplying what I mean, the dispositionalist account of justification is correspondingly unconvincing.

So, is the fact that Jones is disposed to apply 'square' to all and only square objects the very same fact as that he *meant* square by 'square'? Kripke gives any inclination to respond affirmatively here more pause when he writes in the aforesaid summary:

'Computational error, finiteness of my capacity, and other disturbing factors may lead me not to be disposed to respond as I *should*, but if so, I have not acted in accordance with my intentions.'

---

10 Kripke on another occasion states (DT) in a very misleading way: 'What it says is: "'125' is the response you are disposed to give, and (perhaps the reply adds) it would also have been your response in the past". This is misleading because if the dispositionalist does not maintain that the latter counterfactual is true, he cannot be taken to be replying to the sceptic at all.

11 ibid.

12 The point is that if dispositions supply the fact as to whether Jones meant addition or quaddition in the past, there will be a straightforward criterion for whether '125' is justified or not.

13 loc. cit. p.37.
In §5.1, I pointed out that as the sceptical paradox is supposed to be a wholly general problem about meaning and since most predicates do not have infinite extensions or refer to abstract objects, KW ought to start with a predicate such as 'square' and prove his case from there. Finitude or non-finitude of a predicate's extension does not lie at the heart of the matter. The sceptic's paradox centres around the problem of interpretation, and not our inability to encode each of infinitely many instances of a predicate's extension in our brains or minds. So even if all our predicates had finite extensions and we could in principle encode each case, the sceptic's question would still arise: upon exposure to a (temporally) new situation in which I judge 'square' to be applicable, can I be sure that I've got it right? Perhaps the rule for using 'square' was to call all square objects up until now 'square'; henceforth all red wooden objects 'square'; so tomorrow it would be a linguistic infelicity to not call my son's toy train 'square'.

The significance of considerations of finitude therefore resides in the fact that I have in the past (and only ever will have in the future) applied the predicate in question on a finite number of occasions.

Related to this fundamental problem of the finiteness of my linguistic experience, is the problem of error: I might be disposed to make mistakes and surely will for addition problems involving large enough numbers, so how can I be sure that my past disposition was to compute the addition function rather than the subaddition function (which corresponds exactly to my error-ridden arithmetical performance)? These problems certainly make the dispositionalist's task very difficult, but perhaps it is not impossible. The dispositionalist has to show that the fact of my being disposed in the past to use '+' or 'square' in a certain way just is the fact of my meaning
addition then. For reasons that will soon emerge, I think the most plausible version of (DT) is a physicalistic one.

The theory starts from the Chomskyean claim that the fact that people actually do manage to achieve agreement in their responses on the basis of acquaintance with a small number of examples and instructions concerning those examples is strong evidence for their being a common program which they all run. Although logically consistent with an infinite number of functions, the same kind of finite evidence involving additions of numbers as a matter of fact leads us all to the same function. The behavioural evidence of different people at the same time and the same person at different times emitting the same responses to questions involving addition or questions regarding the applicability of 'square' strongly confirms the existence of neurophysiological facts governing our use of these - and other - predicates. But just because there is an inductive leap from the observation of finite convergent behavioural responses to the positing of a specific common rule, we must invoke neurophysiological hypotheses. The sceptic asks how I know that I didn't mean quus by '+' in the past. We can start to answer this, the dispositionalist claims, by first bringing in a neuroscientist who opens up my skull and builds a Turing Machine which replicates its workings. For each predicate whose application I take to be rule-governed, the scientist identifies the distinct states of the TM and the causal mechanism of the particular program corresponding to it. He watches to see whether my hardwiring changes over time with respect to the program I run.

Now we can let the game begin. The sceptic asks his question: 'How do you know that you meant plus by "+" in past?' The dispositionalist asks whether we are to take '+' to mean addition now. The sceptic agrees that we are. The brain scientist then reports that with respect to my '+' program, my TM is
presently in the machine state MS+. We then check the records to see what state the TM was in in the past when I performed addition problems. Again we see that it was only ever in the state MS+. This state seems to be the TM state characteristic of ‘+’ both in the past and now. But the sceptic agreed that I mean plus by ‘+’ now, so, having established that my TM state has not changed, the dispositionalist argues that since my TM state at any time is definitive of what I mean at that time, the sceptic ought to agree that I always meant plus by ‘+’ in the past too. Hence my present response of ‘125’ is justified since I have always meant and mean now, plus by ‘+’.

Kripke, in anticipating this reply, dismisses it as all but worthless:

"Am I supposed to justify my present belief that I meant addition, not quaddition, and hence should answer ‘125’, in terms of a hypothesis about my past dispositions? (Do I record and investigate the past physiology of my brain?) Why am I so sure that one particular hypothesis of this kind is correct, when all my past thoughts can be construed either so that I meant plus or so that I meant quus? Alternatively is the hypothesis to refer to my present dispositions alone, which would hence give the right answer by definition?"

Clearly only my past dispositions are relevant on (DT) to determining what I meant in the past and the dispositionalist’s answer to Kripke’s incredulous question is:

"Yes! I am supposed to justify my present belief that I meant addition in terms of a hypothesis about my past dispositions (how else could I justify a belief about what I meant in the past, if meaning is constituted by dispositions, if not by hypothesising about what my past dispositions actually were?) and yes, the best way to do this is to investigate the physiological basis for the addition program in my brain."

One of Kripke’s worries seems to be that we could never have evidence that the program the TM was running was plus rather than quus. We shall shortly see whether this is well-founded.

But it is clear that Kripke’s main worry with (DT) is that it ignores the question of normativeness - this is why he thinks that the dispositionalist response just could not be right. Thus he says:

14 loc. cit. p.23
'A candidate for what constitutes the state of my meaning one function, rather than another, by a given function sign, ought to be such that, whatever in fact I (am disposed to) do, there is a unique thing that I should do.'

The last sentence is elliptical for something like: 'there is a unique thing I should do if I'm to remain faithful to what I meant by that function sign in the past.' But clarifying the sentence in this manner reveals no apparent disagreement with the dispositionalist for he will concur that one should say '125' if one is in the physical state of intending to accord with what one meant in the past by '+' (if in fact one did mean plus by '+' in the past).

Kripke does not spell out why he thinks the dispositionalist is committed to denying (or, worse, ignoring) this truism.

I turn now to the question of whether, on the version of (DT) sketched before, we can have evidence that I meant addition rather than quaddition by '+' in the past. We should note that the dispositionalist has not produced any evidence that the program I'm running is the addition program; all he has done so far is to argue that whatever I now mean by '+', we can have neurophysiological evidence to suggest I meant exactly the same by it in the past and so, on the assumption that I now mean addition, the evidence suggests that I meant addition before. But there are an infinite number of functions which are consistent with all my actual dispositions up to the present yet diverge in the future, so how can we be sure that this assumption is correct?

The dispositionalist will claim that we have good inductive evidence for believing that my dispositions apropos '+' will remain stable in the future - indeed, the evidence that others are all in the TM state MS' will be good evidence for believing that addition, though only one amongst the infinitely

\[15\] loc. cit. p.24
many logically possible functions that an omniscient being might choose to
associate with the ' + ' functor is the only humanly salient function - the only
one that humans can rationally come to associate with '+'. The dispositionalist
will most probably seek to establish that the function that my TM was
computing was addition through considerations of simplicity. Kripke argues
that such considerations are of no avail in attempting to refute the sceptic
since the sceptic does not understand the competing hypotheses.16 But in the
case of hypotheses about dispositions, Kripke cannot make this complaint.

So would the dispositionalist have established that I meant addition by '+' in
the past if he could choose between alternative hypotheses as to what I
mean on the basis of simplicity considerations? I contend that he would not
have, for there is a crucial premise in his argument which is open to the
sceptic's arguments: that premise is that the TM state is, at any time, to be
definitive of what I mean. Kripke points out that the intention to take the
TM as authoritative is itself open to the paradox17: thus if I say 'I will take
whatever response the TM emits as correct', the sceptic might ask how I
know that I meant correct by 'correct' rather than quorrect (where a
quorrect response ... ).

Is this move justified? Perhaps the dispositionalist does not need to form a
general intention of this type, perhaps it suffices to answer the sceptic if at
each time the dispositionalist endorses TM's output as authoritative of what

---

16 The sceptic argues that there is no fact as to what I meant, whether plus or quus. Now simplicity considerations can help us decide between competing hypotheses, but they obviously can never tell us what the competing hypotheses are. If we do not understand what two hypotheses state, what does it mean to say that one is "more probable" because it is "simpler"? If the two hypotheses are not genuine hypotheses, not assertions of genuine matters of fact, no simplicity considerations will make them so; loc. cit. p.38.

17 loc. cit. p.33. Kripke discusses the attempt to make a physical machine or its program as definitive of what I mean at pp.33–35.
he means then - the intentions might be in sensu diviso rather than in sensu composito. Then, since the sceptic is assuming at any given time that the dispositivealist then means what he thinks and says he means, the content of such a demonstrative intention would be outside of the reach of the paradox?

I do not believe that this will do, but before we see why, we need to draw attention to the fact that the TM in question is a physical machine, since Kripke claims that the dispositivealist cannot take a physical machine as authoritative of my intentions, but must instead choose the TM program. I'm taking it that the TM could just be my brain. Kripke raises several problems for this suggestion. The first is that I can only claim the TM embodies the addition function if I have a means of deciphering the machine language instructions in terms of which the addition program is encoded in the TM, 'the sceptic, of course, being free to interpret all these instructions in a non-standard "quus-like" way'. This gets things backwards: the dispositivealist is not claiming to know in advance of building the machine which function I'm running, the function I encode is to be ascertained from TM's sequence of machine states; certainly any set of instructions formulating my '+'-program invites sceptical reinterpretation ('move left' as 'move right' etc), but a formulation is unnecessary, the values of the function are just given by TM's outputs.

The second is that the TM is finite:

'Indefinitely many programs extend the actual finite behaviour of the machine'. Kripke again puts this point in terms of the finite capacity of the machine - some numbers

---

18 At pp.34-35
19 loc. cit.
20 loc. cit. p.34
'are simply too big' but we can put the point alternatively in terms of the finite computational experience of the TM. Thus expressed, it is not clear to me that there is anything to be gained in moving to the program for the dispositionalist (it has only been and ever will have been run for a finite period of time etc). The third criticism is that since machines can malfunction and since errors can only be identified as such if we know what function the TM is computing, we cannot take the physical machine as constitutive of my intentions, we must choose the program. Why is it impossible in principle for our neuroscientist to produce a detailed enough physical description of the physical workings of TM to exhaustively specify all those states which are physically accessible for a normal running machine? A state that isn't listed a fortiori corresponds to a malfunctioning of the TM.

We can now return to the question raised before: whether, if a physical TM is to be authoritative of my intentions, I need form the general intention across time to take whatever the TM spits out at a time as correct at that time or whether at each time I can simply point to its particular output as what I intend as correct then. It seems that the latter intention is inadequate as it stands because it cannot explain why I might reject an answer as

21 loc. cit.
22 Thus, suppose there are only finitely many numbats in existence and they are all housed in Taronga Park Zoo. I wheel in my TM to the zoo and ask him whether the twelve shy marsupials he observes before him are all numbats - he checks his list, sees that these are all (and only) the beastsies on it and says 'Yes!'. But then my sceptical friend asks him how he knows that he meant numbats rather than numbskulls by 'numbats', where 'numbskulls' are numbats before now, skunks or gulls now and beyond.
23 loc. cit. pp.34-35
24 Here I am only questioning Kripke's reasons for thinking the dispositionalist must choose the program rather than the machine as definitive of my intentions - it is a separate question whether in the face of the fact that I am disposed to err either a machine or a program can tell me what I mean.
incorrect - the machine might spit out '5' as the answer to '68 + 577' and I might suspect that the instruction to print out '125' somehow misfired. So the dispositionalist might say that the intention is really to at each stage accept what the TM says, on the proviso that it is functioning normally (which we are taking to be physically decidable).

Even so, this still will not do. Pace Kripke, for the dispositionalist there is a unique response I should make at any stage if I'm to accord with what I meant by '+' or 'square' in the past, but the problem is that I might have no cognisance of what that response might be and so there can be no sense in which my future self recognises in its own responses then a fulfilment of the dictates of my present self concerning '+' or 'square', unless I have a suitably general intention directed toward the future.

So the dispositionalist must concede that I form a general intention across time to take TM's outputs as authoritative of what I mean, or, at the very least, to augment the de re intentions with an intention to use '+' or 'square' in the same way as before, if he is to explain the sense in which I can intend my future performances to keep faith with my present understanding of these predicates ... and these general intentions just raise the problem anew.

Furthermore, even if the content of such an intention were immune to sceptical worries, the intention to take TM as authoritative of what I mean seems misguided. For it is surely possible that, at a given time, I might classify an object as 'square' and my TM not do so and then, if the present version of (DT) is correct, if we persist after investigation with our disparate evaluations, I would have to accept TM's evaluation as definitive of what I really meant by 'square'.

But this seems intuitively wrong - I seem to understand what I mean by 'square' well enough to make the envisaged stalemate one in which I can be intelligibly thought of as disagreeing with my TM that the object is really square. I am not uncertain whether it is square or not; rather, I am certain that it either is or isn't, my TM to the contrary. So it seems as if I can make the judgement that a TM or even my own brain is in error and so it appears that even if the dispositionalist can show that my dispositions have remained constant over time, all we have is what my brain would give; and since it is possible that my brain does not give me what I mean, dispositions cannot be definitive of meaning.

§5.3.3 Intentions and meaning

I turn now to the second type of answer to the sceptic mentioned in §5.1 - the thesis that to be in the state of intending to use '+' to mean addition is to be in an irreducible mental state. This thesis is apt to look mysterious (as all irreducibility theses are), but I think that there is evidence for believing it to be true. The Intentionalist, as we shall call him, will take the failure of (DT) to show that being disposed to call square things 'square' cannot be what having the intention to so use 'square' as to mean square consists in. Kripke discusses this suggestion at p.51. There, he appraises the suggestion that meaning addition by '+' is:

'simply a primitive state, not to be assimilated to sensations ... or any "qualitative" states, nor to be assimilated to dispositions, but a state of a unique kind of its own' as 'desperate'

since 'it leaves the nature of the postulated primitive state ... completely mysterious.' But more importantly, he claims that Kw has shown:

\[\text{loc. cit.}\]
'that it is logically impossible (or at least there is a considerable logical difficulty) for there to be a state of meaning addition by "plus" at all.‘

Why is it 'logically impossible'? Kripke gives as his reasons that:

'Such a state would have to be a finite object, contained in our finite minds. It does not consist in my explicitly thinking of each case of the addition table, nor even of my encoding each separate case in the brain: we lack the capacity for that ... can we conceive of a finite state which could not be interpreted in a quasi-like way?'

The argument here is very compressed but I think it is supposed to go like this:

(1) Because our linguistic experience with '+' or 'square' is finite (we only ever applied these predicates in finitely many contexts in the past), then any mental state posited as causally explanatory of such finite use must also be, in the favoured sense, finite (i.e. only capable of ever generating finitely many applications in finitely many contexts).

(2) Although it might look as if we can evade this conclusion by stipulating that the mental state has an unbounded generative capacity built into it by virtue of a wholly general content, it is precisely this stipulation which is the sceptic’s target ('Can we conceive of a finite state which could not be interpreted in a quasi-like way?').

(3) So since there are indefinitely many occasions on which we might use '+' or 'square' and since we cannot somehow encode all the different types of case ('we lack the capacity for that') and since the sceptic has shown the idea that my intentions concerning '+' or 'square' might be general in content to be vacuous, the proposed mental state fails, by virtue of its finitude, to provide the sought-for fact.

This argument is flawed. For Kripke has just assumed that the sceptic’s challenge cannot be met - i.e. that the notion of having a general intention such as to so use 'square' as to mean square has been proved by the sceptic to be incoherent.

Wright notes that the sceptic presupposes that one's knowledge of what one means by an expression must be non-inferential. But the demand that knowledge of what I formerly meant by 'square' be derived from knowledge of some other sort of fact, e.g. about my behavioural dispositions, is justifiable only if knowledge of what I presently mean has also to be

26 loc. cit.
27 loc. cit.
inferential: otherwise the sceptic will be answered as soon as I recall what I then meant.

But can we really defend the view that knowledge of present meaning is both non-inferential and sufficiently general in content? What we require, in Wright's words, is:

'a mental state which each of us knows about, at least in his own case, non-inferentially and yet which is infinitely fecund possessing specific directive content for no end of distinct situations.' 29

Wright's claim is that both crucial features - non-inferentiality and indefinite fecundity - "are characteristic of our standard intuitive notion of intention" 30. As to the first feature, Wright's contention is that:

"To come to know you have a certain intention is not to have it dawn on you that you have an intention of some sort and then to recover an account of what the intention is by reflecting upon recent or accompanying thoughts. It is the other way round: you recognise thoughts as specifying the content of an intention you (already) have because you know what the intention is an intention to do." 31

But what of the crucial generality or 'fecundity' as Wright calls it? Wright gives as his example here an intention to prosecute at the earliest possible date anyone who trespasses on my land and points out that:

"there can be no end of distinct responses in distinct situations which I must make if I remember this intention, continue to wish to fulfil it, and correctly apprehend the prevailing circumstances." 32

Is Wright's argument acceptable here or does it beg the question against the sceptic? Although not question-begging, I do not believe that it will do as it stands. Wright stresses that these are the features of 'our intuitive conception of intention', at the same time admitting that it is possible that this conception is radically incoherent 33. 'But', he claims,

29 loc. cit. p.775
30 loc. cit.
31 loc. cit. p.776
32 loc. cit.
33 loc. cit. The sceptic would presumably agree that these are the lineaments of 'our intuitive conception of intention' and go on to argue that it is incoherent!
'the fact is that it is available to confront Kripke's sceptic and that, so far as I can see, the sceptical argument is powerless against it.'

I would have thought that a proof of the coherence of the notion was a sine qua non for the 'confrontation' Wright alludes to to be decided in the non-sceptic's favour. Wright's case could have been established prima facie if the requisite fecundity of intentions was itself guaranteed by their non-inferentiality. But this is not so and so it is open to KW to point out that his paradox attaches to any intention.

Wright's quick dismissal of the sceptic fails - he is certainly correct to point out that if we can show that what I presently mean by '+' or 'square' is both non-inferentially accessible to me and irreducible, then I can answer the sceptic by simply recalling what I previously meant. But what evidence is there for my intentions enjoying such a status?

The best evidence, I think, is that the sceptic's own hypothesis is self-refuting: when we try to articulate how it might be that there just isn't any fact as to what I mean by '+' or 'square', we see that the alleged possibility is a chimera. It cannot be that I mean nothing by '+'. So, the hypothesis that I mean addition and that I have the ability to know that I do is a viable one.

---

34 loc. cit.

35 KW need not deny that it appears to us that when, as we say, we 'form the general intention' to prosecute anyone who trespasses on our land there are no end of distinct responses I will make etc. But his claim is that we misunderstand the appearances - having such an intention cannot be, no matter how much it might appear to be, a matter of somehow encoding the infinitely many directions as to how to respond in each of the infinitely many possible circumstances. KW will claim that 'our intuitive conception of intention' in reality amounts to nothing more than that of having a certain disposition. KW could agree that we have non-inferential access to our dispositions but claim, as before, that my dispositions to use 'square' in a certain way do not obviate the need for me to make a judgement that this is the correct way, in the circumstances, to apply the expression.
- and, contrary to Kripke's Humean deflationary strategy, it does have genuine explanatory power:
It explains how a mathematically gifted Robinson Crusoe, having nutted out the principles of Group Theory for himself, when once he recognises that the Rubik's Cube washed up upon his island's shores can be solved by an application of those principles, suddenly acquires an ability he did not previously possess: not just a brute inclination to be explained away by the familiar Wittgensteinian strategy, for he already had all the brute inclinations he now has before he had mastered the Cube (i.e. to want to get all the colours together etc); what he did not have was the capacity he now possesses: to solve the puzzle, a capacity he could not have acquired in the manner in which he did but for his knowledge of Group Theory. Kripke owes us an explanation of how this latter counterfactual could possibly be true if knowledge of the principles of Group Theory is not causally efficacious - and how could such knowledge be causally efficacious if Robinson Crusoe cannot follow 'the directions he gave himself'?
The best reason for believing that we follow rules and mean what we think we mean by our words is therefore that the contrary supposition is self-refuting and similarly the best reason for thinking that we have non-inferential access to our semantic intentions, which intentions are sui generis, is likewise that Kripke has shown that the alternative ersatz possibilities all fail to amount to adequate notions of meaning.
BIBLIOGRAPHY:

Appiah, A:
(1984) 'Anti-Realism unrealised' Philosophical Quarterly 34
(1985) Assertion and Conditionals

Baldwin, T:
(1975) 'Quantification, modality and indirect speech', in
Blackburn, S. (ed.), Meaning, Reference and Necessity

Belnap, N.
(1962) 'Tonk, Ploak and Plink' Analysis vol. 22, no. 6, 1962,
reprinted in Copi, I. and Gould, J. (eds.) Contemporary Philosophical

Bennett, J.F.
(1973) 'The meaning-nominalist strategy', Foundations of
Language, 10, pp. 141 - 168.
(1976), Linguistic Behaviour Cambridge University Press.
(1985) 'Critical Notice: Inquiries into Truth and Interpretation'

Blackburn, S.
(1984a) Spreading the Word - Groundings in the Philosophy of
Language
pp. 281-301.

Bouveresse, J
and Parret, H (eds)
(1981) Meaning and Understanding

Brandom, R.
LXXIII pp. 137-149.


(1965) 'Theories of meaning and learnable languages' in Bar-Hillel, Y (ed) Logic, Methodology and Philosophy of Science v. Amsterdam, North Holland, 1965 pp. 383-395


Davidson, D. and Hintikka, J. (eds.)
(1975) Words and Objections: Essays on the Work of W.V. Quine,

Davies, M
(1981) Meaning, Quantification and Necessity: Themes in
Philosophical Logic, Routledge and Kegan Paul, Routledge and

Devitt, M

Dummett, M
(1959a) 'Truth' Proceedings of the Aristotlean Society, n.s. vol. 59
(1959b) 'Wittgenstein's Philosophy of Mathematics'
(1969) 'The Reality of the Past' Proceedings of the Aristotlean
(1973b) 'The Justification of Deduction' Proceedings of the British
(1974a) 'The Significance of Quine's Indeterminacy Thesis'
(1975a) The Philosophical Basis of Intuitionistic Logic' in Rose et al.,
(1975b) 'What is a Theory of Meaning?' in Guttenplan, S (ed.), 1975
pp.97-138.
(1976a) 'What is a Theory of Meaning (II)?' in Evans, G. and
Edgington, D.

Evans, G.


Evans, G and McDowell, J (eds.)

Field, H.
Follesdal, D.

Foster, J
(1976) 'Meaning and Truth Theory' in Evans, G and McDowell, J (eds.)

Frege, G
(1952) Translations from the Philosophical Writings of Gottlob Frege
(1952) The Thought: A Logical Enquiry', trans. A. & M. Quinton
(1964) The Basic Laws of Arithmetic, exposition of the system, trans.
(1977) 'Thoughts' in Geach, P (ed) Logical Investigations,

Gabbay, D and Geunthner, F. (eds.)
1986.

Geach, P. T.

Gentzen, G.
(1969) 'Investigations into logical deduction' in Szabo M.E. (ed),
The Collected Papers of Gerhard Gentzen, North-Holland, Amsterdam,
1969.

Grandy, R.
(1973) 'Reference, Meaning and Belief' Journal of Philosophy, 70.
Grice, H. P.


Guttenplan, S. (ed.)

Heyting, A.
(1956) Intuitionism, Amsterdam, North Holland, 1956.

Hintikka, M. and Vermazen, B. (eds.)

Hookway, C. and Pettit, P. (eds.)

Johnston, M.

Kripke, S.

Lear, J.

LePore, E. and McLaughlin, B (eds.)

Lepore, E. (ed.)
Lewis, D. K.


Loar, B.


Lycan, W.

McDowell, J.


McGinn, C.


Margalit, A. (ed.)

Peacocke, C. A. B.


Perry, J


Platts, M. de B.

Prawitz, D.


Prior, A

Putnam, H
(1975) *Mathematics, Matter and Method*, Philosophical Papers vol. 1


Quine, W.V.O

(1966) *The Ways of Paradox and Other Essays*, New York, Random House,
1966.

(1969) *Ontological Relativity and other essays*, Columbia University Press,

(1970) 'On the Reasons for Indeterminacy of Translation', *Journal of


(1975a) 'The Nature of Natural Knowledge' in Guttenplan, S (ed.) 1975,
pp.67-82.

(1975b) 'Mind and Verbal Dispositions' in Guttenplan.

(1977) 'Facts of the Matter' in Shahan, R.W and Merrill, K.R. (eds.)
*American Philosophy from Edwards to Quine*, Norman: University of

pp.90-95.
Rose, H. E. and Shephardson, J. C. (eds.)
(1975) Logic Colloquium '73, Amsterdam, North Holland.

Sainsbury, R. M.

Schiffer, S

Smart, J.J.C.


Strawson, P. F.


Sundholm, G.


Tarski, A.
Taylor, B.


van Dalen, D.

van Straaten, Z. (ed.)

Wiggins, D.

Williams, B.


Wittgenstein, L.
Wright, C.


