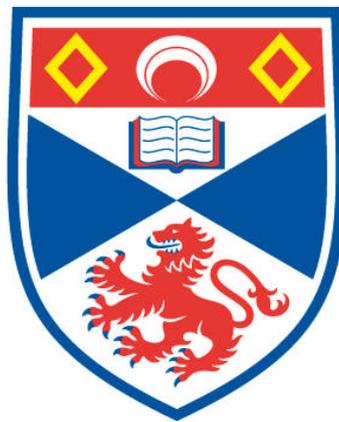


THE SELF AND SELF-CONSCIOUSNESS

Andrew J. Hamilton

**A Thesis Submitted for the Degree of PhD
at the
University of St Andrews**



1988

**Full metadata for this item is available in
Research@StAndrews:FullText
at:**

<http://research-repository.st-andrews.ac.uk/>

Please use this identifier to cite or link to this item:

<http://hdl.handle.net/10023/2704>

This item is protected by original copyright

THE SELF AND SELF-CONSCIOUSNESS

by

ANDREW J. HAMILTON

M.A., M.Phil.

A thesis submitted to the Faculty of Arts of the University of
St. Andrews in fulfilment of the requirements of the degree of
Doctor of Philosophy.

March 21st. 1987



"The door flew open and X, alias 'Baldy', fell out on the road. We pulled up at once, and then he said 'Did anybody fall out?' or 'Who fell out?' I don't exactly remember the words. When told that Baldy fell out, he said 'Did Baldy fall out? Poor Baldy!'"

(William James, 'The
Principles of Psychology',
1890)

I, Andrew John Hamilton, hereby certify that this thesis, which is approximately 75,000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

date 27.11.87..... signature of candidate

I was admitted as a research student under Ordinance No. 12 on 1.10.82 and as a candidate for the degree of Ph.D. on 1.10.83. The higher study for which this is a record was carried out in the University of St. Andrews between 1982 and 1987.

date 27.11.87..... signature of candidate

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Ph.D. in the University of St. Andrews and that the candidate is qualified to submit this thesis in application for that degree.

date signature of supervisor

In submitting this thesis to the University of St. Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker.

ABSTRACT

It is the aim of this thesis to consider two accounts of 1st-person utterances that are often mistakenly conflated - viz. that involving the 'no-reference' view of 'I', and that of the non-assertoric thesis of avowals. The first account says that in a large range of (roughly) 'psychological' uses, 'I' is not a referring expression; the second, that avowals of 1st-personal 'immediate' experience are primarily 'expressive' and not genuine assertions.

The two views are expressions of what I term 'Trojanism'. This viewpoint constitutes one side of a 'Homeric Opposition in the Metaphysics of Experience', and has been endorsed by Wittgenstein throughout his writings; it has received recent expression in Professor Anscombe's article 'The First Person'. I explore the ideas of these writers in some depth, and consider to what extent they stand up to criticism by such notable 'Greek' contenders as P.F. Strawson and Gareth Evans.

I first give neutral accounts of the key-concepts on which subsequent arguments are based. These are the immunity to error through misidentification (IEM) of certain 1st-person utterances, the guaranteed reference of 'I', avowal, and the Generality Constraint. I consider the close relation of Trojanism to solipsism and behaviourism, and then assess the effectiveness of two arguments for that viewpoint - Anscombe's Tank Argument and the argument from IEM. Though each is appealing, neither is decisive; to assess Trojanism properly we need to look at the non-assertoric thesis of avowals, which alone affords the prospect of a resolution of the really intractable problems of the self generated by Cartesianism.

In the course of the latter assessment I consider the different varieties of avowal, broadening the discussion beyond the over-used example 'I am in pain'. I explore Wittgenstein's notion of 'expression', and discuss how this notion may help to explain the authority a subject possesses on his mental states as expressed in avowals. My conclusion is that an expressive account of avowals can

provide a satisfactory counter to the Cartesian account of authority without our needing recourse to a non-assertoric or even to a non-cognitive thesis.

Discussion of self-consciousness is implicit in discussion of the Homeric Opposition, but there is in addition a short chapter on the concept itself.

ACKNOWLEDGEMENTS

My first and most important debt of gratitude is to my Supervisor, Professor Crispin Wright, without whose tireless support and patient criticism this thesis could not have been produced. There can be few research students who have been so fortunate in their choice of Supervisor. If there is a definite attitude and approach towards philosophical questions displayed herein, it is largely the product of his teaching over a number of years. My appreciation is due also to Roger Squires, who took over supervision temporarily in Professor Wright's absence. His cautions against the dangers of theorising in philosophy have often served to awaken me from my more dogmatic slumberings.

The research of which this thesis is the product has been funded principally by the University of St. Andrews, and also by the Humanitarian Trust, the Sir Richard Stapley Educational Trust and the Gilchrist Foundation. To these my gratitude is due.

Many thanks also to Mary Taylor and Mark Stringer, for typing under some pressure the first and final drafts respectively.

I have received much support, philosophical and moral, from friends and family in the course of the work. Two friends in particular helped at a time when it seemed the thesis would never be completed - Catherine Wright and Anne Currie. I am also very grateful to my mother, whose patient forbearance and support in the face of an incomprehensible and seemingly interminable project have been out of the ordinary.

The thesis is dedicated to the memory of my father.

Andrew Hamilton
St. Andrews.

CONTENTS

	<u>Page</u>
Abstract	iv
Acknowledgements	vi
Introduction	x

CHAPTER ONE

1.1) <u>Immunity to Error through Misidentification (IEM) (I)</u>	3
(i) The literary background	
(ii) Argument to 'no-reference' view of 'I'; neutral characterisation of IEM	
(iii) Baldy's error	
1.2) <u>IEM and Guaranteed Reference</u>	10
(i) Anscombe's conflation	
(ii) Formulation of 'guaranteed reference'	
(iii) Status of the phenomenon	
1.3) <u>Avowal</u>	16
(i) A neutral characterisation	
(ii) The relationship of IEM and avowal	
1.4) <u>IEM (II)</u>	23
(i) Avowals are IEM	
(ii) Are demonstrative judgements IEM?	
1.5) <u>The Generality Constraint</u>	27

CHAPTER TWO

2.1) <u>Historical Account and Definition of 'No-Reference' View - Wittgenstein's 'Philosophical Remarks'</u>	33
(i) What is Trojanism?	
(ii) The 'Philosophical Remarks' account	
(iii) The danger of 'covert behaviourism'	
2.2) <u>Solipsism and the 'No-Reference' View</u>	41
(i) Epistemological and metaphysical solipsism	
(ii) The connection between Trojanism and solipsism	
(iii) Wittgenstein on 'what the solipsist wants'	
2.3) <u>Arguments against Trojanism: from the Generality Constraint et al.</u>	49
(i) Militant and moderate Trojanism and the 'continuous ability' challenge	
(ii) Evans' argument: 1st-person undecidable propositions	
(iii) Trojan response and deficiencies	
2.4) <u>Anscombe's Tank Argument (I): The Self-Reference of 'I'</u>	57
(i) Introductory	
(ii) Qualifying the self-reference principle	
(iii) 'A'- and 'I'-reflexives; circularity	
(iv) Discussion of non-self-conscious self-reference	

2.5)	<u>Anscombe's Tank Argument (II): Presentation of the Argument</u>	64
	(i) Prolegomenon	
	(ii) The Tank Argument of the first part of Anscombe's dialectic	
	(iii) A fallacious false start	
	(iv) Climax of the Anscombian dialectic	
2.6)	<u>Anscombe's Tank Argument (III): Assessment</u>	74
	(i) Three Greek options	
	(ii) Kenny on 'The First Person'	
	(iii) The argument from pragmatic self-defeat	
	(iv) A better line of argument	
2.7)	<u>Trojan Arguments from IEM (I): The Unique Guarantee for 'I'</u>	87
	(i) 'No unnoticed substitution' of the object of self-reference is not the 'unique guarantee', and is empty	
	(ii) The unique IEM guarantee and the substantial Trojan case	
	(iii) Parfitian considerations	
2.8)	<u>Trojan Arguments from IEM (II): The Wittgensteinian Route to the 'No-Reference' View</u>	100
	(i) Two Trojan arguments	
	(ii) 'If you can't be wrong, you can't be right either'	
	(iii) A first Greek response: Identification can be 'passive'	
	(iv) A second Greek response: 'Active' identification can be infallible	
 <u>CHAPTER THREE</u>		
3.1)	<u>Nozick on 'Reflexive Self-Reference'</u>	113
	(i) 'Self-reference' and 'auto-reference'	
	(ii) A muddle over IEM and 'reflexive access'	
	(iii) The self-synthetic 'tour de force'	
3.2)	<u>'Fortuitous' and 'Aware' Self-Reference; the Circularity of the Self-Reference Principle</u>	126
	(i) Anscombe's 'striking phenomenon'	
	(ii) Is the self-reference principle circular?	
	(iii) The consequences of circularity	
3.3)	<u>IEM and Self-Consciousness</u>	136
 <u>CHAPTER FOUR</u>		
4.1)	<u>The Problem of Avowals</u>	143
4.2)	<u>Avowals of Belief</u>	148
	(i) The categories of avowal	
	(ii) Sources of error in avowals of belief: Identification and attitude	
	(iii) Sources of error: Propositional element	
	(iv) Evans' account of 'belief self-ascription'	
	(v) Wittgenstein's account: 'Moore's paradox'	
	(vi) What does Evans' account show? Qualified avowals defined	
4.3)	<u>Avowals of Desire</u>	165

4.4)	<u>Avowals of Perceptual Experience</u>	169
	(i) 'Sense-data' and non-avowals of perceptual experience	
	(ii) 'Avowal' is a category of tokens, not types	
	(iii) Is there 'an interesting class of avowals?'	
	(iv) Evans on self-ascription of perceptual experience; 'intermediate avowals'	
4.5)	<u>An Argument to the Non-Assertoric Thesis</u>	180
	(i) Statement of the argument from non-justifiability	
	(ii) Responses	
	(iii) The 'anti-Cartesian motive' for the thesis	
4.6)	<u>'Expression' (I)</u>	187
	(i) The intuitive notion	
	(ii) Wittgenstein's account	
4.7)	<u>Problems Faced by the Non-Assertoric Thesis</u>	195
	(i) 'Covert behaviourism'	
	(ii) Truth-value links (I): How could avowals subserve inferences?	
	(iii) Truth-value links (II): The Haldane Objection	
4.8)	<u>Do we need a Non-Assertoric Thesis anyway?</u>	204
	(i) 'Non-force theses'	
	(ii) A non-cognitive thesis on its own	
	(iii) Even a non-cognitive thesis is unnecessary	
4.9)	<u>Resolution (I): Knowledge and Authority</u>	214
	(i) Subject and expression	
	(ii) The 'thin' sense of 'knowledge'	
	(iii) Arguing against an 'observational' account of self-knowledge	
4.10)	<u>Resolution (II): 'Expression' (II)</u>	221
	(i) Taylor's account	
	(ii) Amending Taylor's account	
	(iii) Defeasibility of mental-episode ascriptions	
	(iv) Consequences of the amended account of 'expression'	
	(v) Dealing with 'qualified' avowals	
	(vi) Avowals and 'pseudo'-avowals	
	Appendix	235
	References	244

INTRODUCTION

This thesis is an account of a 'Homeric Opposition in the Metaphysics of Experience' intended to rival the Homeric struggle in the philosophy of language outlined by Strawson in his article 'Meaning and Truth' ([1971] p. 172). Whether this is an apt description, or whether a more exact parallel would be with the mock-heroic skirmish of Alexander Pope's 'The Rape of the Lock', is a question for the reader to decide. Perhaps the 'Greek' and 'Trojan' terminology eventually became something of an encumbrance; I am so used to it, it is hard for me to judge. The initial characterisation of these opposed viewpoints was in terms of their (Trojan) endorsement or (Greek) rejection of the proposal that 'I' is, in a large class of uses, not a referring expression, and the self no object. This characterisation still has validity - but as explained in Chapters 2.1 and 4, things turn out not to be so simple.

The provocation for this work on the self came in the first place from Professor Anscombe's article 'The First Person' [1975]. This article constitutes probably the most convincing presentation of the Trojan viewpoint, despite its unclarity about some of the key-concepts in the area. As an expression of the opposed viewpoint, the chapter entitled 'Self-Identification' in Gareth Evans' The Varieties of Reference [1982] is unrivalled. It is the most profound discussion of self-consciousness I have encountered. Underpinning each discussion, but interpreted or responded to differently, are the ideas on the self expressed by Wittgenstein from the period of the Notebooks for the Tractatus to that of the Investigations. It is a major part of the present undertaking to attempt to clarify the frequently cryptic and frustrating remarks in question.

Non-philosophers at least will be sceptical whether an investigation of the semantics of the little word 'I' will yield a lot in the way of understanding of Self-Consciousness. Of course, the present investigation is one almost entirely within Philosophy and not Psychology. And the ramifications of the discussion of the

Homeric Opposition has resulted in the treatment of Self-Consciousness per se, even in its philosophical sense, being limited to a rather short Chapter 3. However, the few explicit connections with psychology that are made would I hope help to illustrate what I believe to be true - that at the end of the road of austere philosophical analysis lies a greater understanding of ourselves.

CHAPTER

ONE

In this first chapter I aim to give, as far as possible, neutral accounts of the key-concepts to be employed in the arguments of subsequent chapters. Neutral, that is, as between Greek and Trojan viewpoints. These concepts are Immunity to Error through Misidentification (IEM) of certain 1st-person utterances, Avowal - and, less importantly, Guaranteed Reference (of 'I') and the Generality Constraint. The idea is that there are facts of all these matters which can be argued on by both sides in the Homeric Opposition. Disagreement arises from the differing interpretations placed on these facts. Hopefully the reader will bear with the technical analysis involved, in the assurance that it will be put to good use later.

1.1) Immunity to Error through Misidentification (IEM) (I)

(i) The literary background

The truism that certain 1st-person utterances are immune to error through misidentification of the subject (a phenomenon referred to henceforth by the abbreviation IEM) was first noticed by Wittgenstein in the Blue Book. In a passage much-quoted in the annals of self-consciousness, he writes:

There are two different cases in the use of the word 'I' (or 'my') which I might call the 'use as object' and 'the use as subject'. Examples of the first kind of use are these: "My arm is broken", "I have grown six inches", "I have a bump on my forehead", "The wind blows my hair about". Examples of the second kind are: "I see so-and-so", "I hear so-and-so", "I try to lift my arm", "I think it will rain", "I have toothache". One can point to the difference between these two categories by saying: The cases of the first category involve the recognition of a particular person, and there is in these cases the possibility of an error or as I should rather put it: The possibility of an error has been provided for ... [but] it is as impossible that in making the statement "I have toothache", I should have mistaken another person for myself, as it is to moan with pain by mistake, having mistaken someone else for me (BB pp. 66-7).

It is the task of the first section of this chapter to refine the principle of IEM as a preliminary to the more substantially argumentative work of Chapters 2 and 3 in which this principle figures. The aim is to produce a characterisation neutral as between the Greek and Trojan viewpoints, and it is worth pointing out initially that Wittgenstein's Trojan sympathies are apparent in the passage above in the 'use as object' / 'use as subject' terminology. ('Use as subject' means 'as subject alone' - there is to be no object in such uses.) Also, the use of 'I have toothache' as an example of IEM is not helpful, since it is one of the sub-class of

IEM-utterances known as avowals, and hence mistaken conflation of IEM and avowal is invited (see pp. 18-21).

What happens when the IEM principle is ignored may be illustrated by two famous literary examples in philosophy. The first is the melancholy death-bed scene in Dickens' Hard Times, where Mrs. Gradgrind is by now drifting away:

'Are you in pain, dear mother?'

'I think there's a pain somewhere in the room', said Mrs. Gradgrind, 'but I couldn't positively say that I have got it.' After this strange speech, she lay silent for some time. (Book II, Chapter 9)

On a lighter note there is from William James the remarkable adventure of one 'Baldy', quoted by Anscombe in her [1975] (pp. 64-65). In a footnote to his discussion of the feeling of absence of self, James writes:

In half-stunned states self-consciousness may lapse. A friend writes me: "We were driving back from ---- in a wagonette. The door flew open and X, alias 'Baldy', fell out on the road. We pulled up at once, and then he said 'Did anybody fall out?' or 'Who fell out?' - I don't exactly remember the words. When told that Baldy fell out, he said 'Did Baldy fall out? Poor Baldy!'" ([1890] Vol. I p. 273n).

Since the adventures of Baldy form a leitmotif throughout our discussion of IEM, it is important to diagnose his 'error' correctly from the outset. In fact we must, for present purposes, suppress one interpretation which would be philosophically uninteresting. For it is possible that the trauma has induced in Baldy a general amnesia - if his question was, in fact, 'Did anybody fall out?' And he does indeed seem to have forgotten that his name is Baldy, hence his final 3rd-person utterance; though it is an unexplained feature of the incident that he seems to have been told 'Baldy fell out' and not 'You fell out'. It is only on a second interpretation that he commits the 'Baldy-type error' that is of philosophical interest. If indeed he did ask 'Who fell out?', then James has some justification in his diagnosis of a 'lapse of self-consciousness'. This

question is a manifestation not of some bizarre kind of amnesia, but rather of a kind of ignorance or doubt, like that of Mrs. Gradgrind, for which there is no real logical provision.

(ii) Argument to 'no-reference' view of 'I'; neutral characterisation of IEM

The relation between IEM and self-consciousness is a topic for Chapter 3. Of more immediate concern is the idea that the phenomenon of IEM is a ground for the 'no-reference' view of 'I'. The move here is from the impossibility of misidentifying the object that I am (or from the absence of criteria for its identification) to the conclusion that there is no identification of such an object either - that there is in fact no object and 'I' does not refer. Writers who are firmly opposed to the 'no-reference' view see it as important to explain why, on their account, this move is illicit. Strawson attempts to explain why 'I' can be used without criteria of subject-identity and yet refer to a subject because even in such uses, there are links with 3rd-person criteria ([1966] p. 165). Shoemaker writes: 'the absence of [the] possibility [of misidentification] ... is one of the main sources of the mistaken opinion that one cannot be an object to oneself, which in turn is a source of the view that "I" does not refer'. He believes that the source of this mistake is the idea that there can be no other kind of model for the self-awareness that grounds IEM-judgements than the perceptual model, and that since this is a preposterous account, there is (so it would appear) no self-awareness here at all ([1968] pp. 561ff.). Finally, Evans locates the supposed error in an equivocation between identification in the sense in which it involves criteria of recognition etc., and in the sense in which it involves no more than object-direction in thought ([1982] p. 218).

How the argument from IEM to 'no-reference' fares in the face of such criticism will be discussed in the next chapter. What is required first is a quite neutral characterisation of the important phenomenon of IEM. In the course of this characterisation, IEM will be distinguished first from the 'guaranteed reference' of 'I', and then from the phenomenon of avowal. Now, the first thing to note is that it is not brought out in Wittgenstein's account, and is denied

or imperfectly understood in the accounts of Strawson and Shoemaker respectively, that IEM can apply to cases of 'physical' as well as 'mental' self-ascription (Strawson [1966] pp. 164-5, Shoemaker [1968] p. 557). (Shoemaker actually concedes that IEM can obtain in physical self-ascription, but only derivatively; e.g. 'I am facing a table' has IEM only in virtue of 'There is a table in my field of vision' being IEM.) This fact would be less likely to be ignored or denied if a further feature of IEM were noted: that it is particular tokens of 1st-person sentences that display IEM, and not sentence-types (to use the Strawsonian terminology). The principle of IEM has to do with the grounds for particular assertions and not merely with their propositional content. An initial characterisation is as follows:

(IEM₁) On certain occasions of use of 'I fell out of the carriage' etc., the utterance is immune to error through misidentification of someone else as the subject.

Now, let us analyse 1st- and 3rd-person ascriptions in the following way:

$$Fa = (\exists x)(Fx \ \& \ x = a)$$

'Fx' is the predication component of the sentence, and 'x = a' the identification component (following the terminology of Evans in his [1982]). So we would informally analyse the sentence 'I am now undergoing intellectual discomfort' as follows:

(Predication component): Someone is now undergoing intellectual discomfort.

(Identification component): That someone is myself.

This analysis suggests the following:

(IEM₂) On certain occasions of use of 'I fell out of the carriage' etc., any source of error must lie in the predication component.

A further reformulation is required in order to focus on the idea that a certain sort of doubt is impossible, and to make clear that

the IEM and non-IEM uses are distinguished by grounds:

(IEM₃) When based on a certain type of ground, the assertion of 'I fell out of the carriage' is IEM iff the ground is such that:

a doubt about the assertion must enjoin the thought that there's no reason to believe that anyone fell out.

(iii) Baldy's error

We can now state Baldy's error a little more precisely than Anscombe does (op. cit. pp. 64-65). To the non-philosophic eye, it might appear that the character of Baldy's knowledge of the mishap is the same as that of one of his fellow passengers in the wagonette, a Mr. Ballard, a deaf-mute with a penchant for rural rides (James [1890] pp. 266-9 and Wittgenstein PI 342). In his semi-autobiographical work "A Queer Memory Phenomenon" and other Recollections, Ballard writes:

I was engaged in one of my well-known reveries on the topic of cosmology and the Divine Providence when I was aroused by the carriage jolting. The door was flung open and a figure disappeared from view. The wagonette was carrying a full load that afternoon including some persons to whom I had not been introduced, and I could not positively identify the poor fellow. There was a frightful commotion with people shouting but of course I could hear nothing. My dog 'Russell' began to simulate pain, as he is wont to do on such occasions. Mrs. Pennyman was quite distressed but managed to intimate to me that the fellow who had fallen from the carriage was X, alias 'Baldy'.

Both Baldy and Mr. Ballard (before the moment of intimation) are aware that someone fell out of the carriage (or so we are tempted to say with Baldy). The content of their knowledge is the predication component alone of 'Baldy fell out of the carriage' / 'I fell out of the carriage', viz. 'Someone fell out of the carriage'. Their attitude to the thought 'Baldy was that someone' (or additionally, in the case of Baldy, 'I was that someone') will be agnostic. Of course, with Mr. Ballard, this is logically quite permissible.

There is no ground for this 3rd-person assertion such that, when based on that ground, the assertion exhibits IEM. But there is, in the case of Baldy, an implicit IEM-utterance the identification component of which he is doubting - nonsensically. To put this more satisfactorily requires a corollary of (IEM₃):

(IEM₄) When based on a certain type of ground (one which includes a ground satisfying (IEM₃)), X's knowledge that someone fell out of the carriage, when it is true that he himself fell out, is such that:

X knows that someone fell out
 = X knows that he himself ('I'-reflexive) is that
 person.¹

When I say that the ground includes that satisfying (IEM₃), I mean that it consists in the latter plus the rule for inferring 'Someone fell out' from 'I fell out'. What is so perplexing about Baldy's case is that his knowledge that someone fell out of the carriage has to be grounded derivatively on knowledge of something of which he appears ignorant - that he himself fell out. This intolerable state of affairs compels us to say that we can give no account of Baldy's knowledge of the incident, but that he certainly does not have knowledge he appeared to have.

It is not always the case that, when it is true that X has fallen out of the carriage, his knowledge that someone fell out has to be grounded in this way. It is by analogy with these other cases that it may appear possible to us that Baldy knows, and knows only, what he appears to know. The events I now relate went unrecorded by William James; they occurred after the publication of his Principles of Psychology and pertain to that period when X alias 'Baldy' had achieved some notoriety as a psychiatric case-study. I quote from the inordinately lengthy serialisation of his Adventures in the Baltimore Oriole [1895]:

I awoke in hospital with no recollection of why I was there, nor indeed of past events in my life at all, except for a curious

¹ For explanation of the 'I'-reflexive, see p. 60

sense of déjà vu. I had been found wandering around the nearby village unable even to say what my name was. On the second day in the hospital I overheard one of the nurses saying that there had been a strange discovery in the next village. A horse had been found wandering in a field, harnessed to a wagonette. Its driver must have fallen out but was nowhere to be found. I thought no more of this incident until later that day a nurse came up and presented me with a much-thumbed copy of William James' Principles of Psychology, on the inside cover of which was written 'Ex libris X alias "Baldy"'. 'Is this familiar to you?' she asked. 'It was found in an empty wagonette which a riderless horse was dragging across a field near Jamesville'. It then struck me in a blinding flash: 'I am Baldy! It was I who fell out of the wagonette!' And the details of my past life came flooding back in all their multifarious richness

The philosophical point is clear despite the regrettably tedious and banal prose style which the writer adopts. The kind of ground afforded by testimony for the assertion 'Someone fell out' is such that the subsequent discovery 'I am that someone' is possible. Here knowledge of 'I fell out' can be grounded derivatively on knowledge of 'Someone fell out'; in the earlier example, the derivation must be the reverse. In the later adventure, 'I fell out of the wagonette' is a non-IEM use.

How in general might the grounds for IEM- and non-IEM utterances be distinguished? McGinn, in his [1983] pp. 51-2, states, without arguing for it, that each IEM-utterance is grounded in awareness of what I refer to as an avowable-state (viz. a mental state of the sort that can be expressed by means of an avowal, that avowal affording evidence for its obtaining which given certain conditions is decisive). This is a tempting idea which would serve to provide a clear account of that which so far has been presented solely by illustration: viz. the way some grounds generate IEM-, and others non-IEM, utterances. I postpone offering an account of this idea until after the neutral characterisation of avowals.

1.2) IEM and Guaranteed Reference(i) Anscombe's conflation

There is a third phenomenon associated with the 1st-person which is distinct from IEM and avowal and which, like IEM, is involved in an argument to the 'no-reference' view. This is the 'guaranteed reference' of 'I', which figures in Anscombe's Tank Argument and which needs to be separated from IEM. There are various features that go under this name. In the formulations that follow I refer, for the sake of economy, to only two of these - the ideas that in using 'I', one cannot fail to refer to anything, and that one cannot fail to refer to that to which one intends to refer. (Other phenomena that could be mentioned in addition include that of 'no unnoticed substitution' - that in using 'I', one refers to the same object on every occasion of use.)¹ Anscombe notices these two features but seems to confuse the second with IEM:

We saw that the 'A'-user would not be immune to mistaken identification of someone else as 'A'. Will it also be so with 'I'?²

The suggestion seems absurd ... if 'I' is a 'referring expression' at all, it has both kinds of guaranteed reference. The object an 'I'-user means by it must exist so long as he is using 'I', nor can he take the wrong object to be the object he means by 'I'. (The bishop may take the lady's knee for his, but could he take the lady herself to be himself?) ([1975] p. 123).

¹ See pp. 87-90.

² 'A' is a term used by Anscombe and functioning as a self-referential device which is not expressive of self-consciousness (see pp. 136-39).

Now it is necessary to defer explicit discussion of the proviso 'If "I" is a referring expression', to deal briefly with this confusion. We have seen that the 'suggestion' that some 1st-person utterances are not immune to error through misidentification is not 'absurd', but correct. In these uses, 'I' does resemble 'A'.

An example of a non-IEM 1st-person judgement might be the following: Gentle John and other heavyweights are being weighed on weighing machines, but for no good reason someone has wired up meter A to machine C, B to D etc. Gentle John says 'I weigh 25 stone'. But his judgement is false because it is Wee Willie who weighs 25 stone, Gentle John in fact clocking in at a mere 23 stone. We then tell Gentle John that someone has tampered with the weighing machines, and that some of the meters are in fact connected to the wrong machines. It then makes sense for the gentle giant to say: 'Someone weighs 25 stone, but is it myself who weighs 25 stone? Maybe it's Wee Willie, or Fat George' So, in the terminology of (IEM)₂, the assertion is doubted due to doubt about the RHS - the identification component. A similar example could be constructed from the further adventure of Baldy, if it turned out that it was not Baldy who had fallen out of the wagonette. But in all such cases the possibility of error due to misidentification requires that the speaker refers to himself. (If 'I' is a referring expression.) And he must always refer to himself when he comprehendingly uses 'I', if the IEM / non-IEM categorisation of 'I'-uses is to be exhaustive. For if there wasn't this guaranteed reference in the second sense (viz. that the speaker refers to that to which he intends to refer, himself), there would be a third category of 'mutant' IEM-utterances. In these cases, I could not misidentify someone else as myself, because I would not be making any claims about myself at all, and so could not make that kind of mistake. But this is all so much nonsense.

To return to the quotation from Anscombe, then, Peacocke (in his [1983] p. 151) is clearly right to say that not only can the bishop take the lady's knee for his, but he can also take the lady herself to be himself - if he saw a woman dressed as a bishop in a mirror, for instance. He might then say 'How splendid I look!' which would be erroneous (or at least ill-grounded) through

misidentification. (It is a non-IEM utterance.) What Anscombe should be saying is not that mistaken identification is impossible in the 1st-person, but that mistaken reference is. (Perhaps it is strange to talk of identifying or misidentifying oneself in a mirror, but I think the idea is clear.)

(ii) Formulation of 'guaranteed reference'

How should the 'guaranteed reference' of 'I' be formulated, then? It will be recalled that at this stage an account that is neutral between the Greek and Trojan views is required, but the account of the phenomenon so far has not been careful enough in this regard. Anscombe points us in the right direction. She adduces various features of the use of 'I' which support the following contention: If 'I' refers, it has guaranteed reference. Our mistake, she claims, is to treat these features as if they established directly that 'I' has guaranteed reference; but they establish only the conditional (and, given the incoherence of 'guaranteed reference' we must negate the antecedent, she claims). The features are those that comprise the idea that there is no sense, in the case of 'I', to the supposition of mistaken reference. Now it is difficult to describe these features without presupposing that one does in fact use a referring expression 'I' - a problem for Anscombe! The features require care in their presentation. Unfortunately this care is not bestowed by Anscombe. She writes:

Getting hold of the wrong object is excluded, and that makes us think that getting hold of the right object is guaranteed. But the reason is that there is no getting hold of an object at all. (ibid. p. 59).

But if we can say categorically that getting hold of the wrong object is excluded, we can also say that getting hold of no object is excluded; incompatible with the Anscombe diagnosis that there is no getting hold of an object at all! No, what we need is some such formulation to govern the principal features as the following:

(G) In using 'I' one cannot be said both to attempt an act
 1 of reference and yet either fail to refer to anything, or

fail to refer to that to which one intends to refer.

Though further refinement is required to limit the uses of 'I' under consideration to those which are comprehending and serious:

- (G₂) In using 'I' one cannot be said both to (i) intend to refer in that way which if 'I' is a referring expression, constitutes serious, comprehending use, and (ii) fail to refer to anything, or fail to refer to that to which one intends to refer.

This tortuous way of putting things means that, as an alternative to (i) and (ii), perhaps no act of reference is involved in the use of 'I', and its serious, comprehending use is characterised in some other way than it is in (i). The characterisation is thus a neutral one (by dint of a certain artificiality!). It is claimed that if 'I' refers, it has guaranteed reference; but the possibility is left open that it does not refer. (Our characterisation of IEM makes no claim about whether or not 'I' refers. Further, there is no need with IEM for a clause about serious and comprehending use, since this is implied by a characterisation in terms of an assertion with a certain ground - see p. 7.)

The question arises whether 'guaranteed reference' is a feature of other referring terms apart from 'I'; or whether it may not in any case be an empty, purely formal feature (in contrast to IEM). The latter in fact seems to be the case. (I shall avoid complication at this point by assuming 'I' does refer; the reader may guess the tedium of continuing in the previous neutral vein, whilst being assured that it could be done.) For guaranteed reference, in common with 'no unnoticed substitution' (see below pp. 87-90), is generated by one's way of describing the situation. There is a condition on the use of 'I' for it to exhibit guaranteed reference: that the use be comprehending. With this in mind, let us look at the two aspects of guaranteed reference specified by Anscombe:

- (1) The object the 'I'-user intends to refer to must exist;
- (2) The intended and the actual referents must coincide.

Now, what must the speaker comprehend about 'I'? For a start, that an 'I'-user refers to himself. The referent therefore must exist for a comprehending 'I'-utterance to be possible. The speaker must also comprehend that an 'I'-user intends to refer to himself. So if the use of 'I' is comprehending, the intended and actual referents will coincide. Built into our description of the use of 'I' as comprehending, will lie the Anscombian 'guaranteed reference'.

(iii) Status of the phenomenon

Is guaranteed reference shared with proper names? I see a figure, which I take to be Bernard playing football at the other end of the beach. I say 'Bernard is playing football'. But it turns out on closer inspection to be someone else. Now a different (though familiar) example. I see someone's reflection in a shop window. I say, on the basis of this, 'My hair is blowing in the wind' (maybe, like Oliver Sacks' 'Disembodied Lady', my proprioceptive sense is impaired¹). But the reflection turns out to be of someone else. In each example, the semantic reference is clear - it is to Bernard, and to myself respectively (this is why the statements are false). The speaker's reference is perhaps not so clear. Presumably it is to the person playing football, and to the person reflected in the window, respectively. But there may be some temptation to say that since the speaker intends to refer to himself, in the second case, it is he who is the speaker's referent too - but the very same consideration could be adduced in favour of regarding Bernard as the speaker's referent in the other case.

The afore-mentioned considerations indicate that the guaranteed reference of 'I' is not a very interesting phenomenon.² In non-IEM cases, there is no greater reason to assert a necessary coincidence of speaker's and semantic reference than in the case of proper names. And there is as much reason to claim that

¹ See his [1985] pp. 42-52.

² Guaranteed reference nonetheless has a role to play in the Tank Argument (Ch. 2.4). This fact points to an unresolved tension.

intended and actual referents coincide in the case of proper names, as in the case of 'I': further, this coincidence appears to be a product of the requirement of comprehending use. There is of course a divergence between the behaviour of proper names and of IEM uses of 'I'; but it would be simply confusing to call this guaranteed reference of 'I', since it is a guarantee about correct identification that marks the difference.

1.3) Avowal(i) A neutral characterisation

Avowals, in the technical sense in which the term is used, are a sub-class of 1st-person, present-tense psychological utterances which, when uttered comprehendingly and sincerely, have the property of guaranteeing the truth of the corresponding 3rd-person ascription. Examples would include 'I'm angry', 'It hurts!', 'I am very excited', 'I am in pain' (of course), 'I'm hot', 'I'm miserable', 'I feel rather cheerful', etc. These, at least, are what are later termed 'paradigm' avowals (of sensation, certain moods and feelings, etc.); complications, postponed till Chapter 4, arise with avowals of propositional attitude, perceptual experience etc. Crispin Wright has suggested five salient features of avowals which form the basis for the following neutral account of the phenomenon. These are intended simply as 'salient features'; it may be that they are not independent one of another, but for present purposes it is sufficient that they are neutral between the Greek and Trojan viewpoints. 'I am in pain' here figures as a 'typical' avowal:

- (i) There are no grounds for the utterance.
- (ii) No misidentification of a subject is possible.
- (iii) My very utterance of 'I am in pain' is a criterion for the correctness of 'He is in pain' said of myself by another.
- (iv) When X, comprehendingly and without intending to deceive, utters 'I am in pain', the truth of 'He is in pain' said of X is guaranteed. Furthermore, except where there is independent evidence that the utterance is not made comprehendingly, when X utters 'I am in pain', the extent of my confidence in the truth of 'X is in pain' is exactly proportional to that of my confidence in the truth of 'X comprehendingly and without intending to deceive avowed "I am in pain"'.

(v) 'I doubt whether I am in pain' has no use.

These features I term the Famous Five. (A modified version of the Five as they pertain to the more 'complicated' cases - 'qualified' avowals - is presented on p. 161.)

Care has been taken in the characterisation to give an account of avowal acceptable to both sides of the Homeric Opposition. Comment on the sense, if any, in which avowals involve identification of a subject, is avoided by referring to the impossibility of misidentification. Talk of the 'truth' or 'correctness' of an avowal is eschewed. Instead of saying that an avowal expresses logically infallible knowledge for the subject (a notion regarded with disfavour by Trojans), the matter is formulated in terms of the senselessness of doubt. Feature (iv) (second part) expresses the fact that if the speaker is taken to understand what he is saying then evidence for or against the non-deceiving nature of his utterance is simultaneously evidence for or against the truth of the corresponding 3rd-person ascription. This relation holds because there is no room for an 'honest mistake' by the speaker - the possibility of non-deceiving but mistaken utterance is ruled out (hence Feature (v)). So evidence against the speaker's intention to deceive rises and falls with that for the 3rd-person ascription (if we assume a comprehending speaker). Take the regrettably widespread phenomenon of Schadenfreude (pleasure in the misfortunes of others). Unfortunate X relates his recent misadventure to envious Y, who fails to suppress completely his wicked smile of satisfaction. 'I'm not sure you're not rather pleased at what has happened', X complains. 'My dear fellow, I could not be more distressed at your misfortune', protests Y. The obvious point is that there will be no evidence that Y is pleased, that will not also be evidence that his disavowal was insincere. (This is to take it that Y's utterance is, in our technical sense, an 'avowal', and that Y's wickedness is not sustained by any self-delusion of virtue. To talk of self-deception, or at least of self-ignorance, in such cases, is to admit the possibility of one's being honestly mistaken in the self-ascriptive application of the very same predicate which on other occasions of use figures in self-ascriptions about which doubt is

inappropriate. The connection with truth which is present in the case of deception is absent in the case of lack of understanding, so that there is the possibility of the kind of 'independent' evidence referred to in Feature (iv) in the latter case but not the former. This evidence sustains the charge e.g. 'X's avowal was not a comprehending one' whilst not committing one to e.g. 'X was not in pain'. Thus the following improbable kind of counter-example is possible. Roberto, a known Spanish monoglot, drops a heavy weight on his foot, by strange coincidence he utters the words 'I am in pain', which he has heard on a 'Let's speak English!' tape, and hops around clutching his foot. Here, I can be completely confident that Roberto is in pain, whilst having no confidence that his was a comprehending avowal.¹ Obviously it is the presence of non-linguistic behavioural criteria which makes such divergent ascription possible, and which grounds any belief that the speaker intends to deceive. Finally, strictly speaking we should use the inelegant locution 'without intending to deceive', instead of 'sincerely', because it is possible that sincere intentions will have no role to play in a non-assertoric thesis - if, for instance, avowals are assimilated to purely expressive behaviour. In the latter case, the fact that an avowal is not intended to deceive will not imply that it was uttered with sincere intention, since the avowal may not be an intentional action at all.

(ii) The relationship of IEM and avowal

What is the correct relationship between the phenomena of IEM and avowal? I wrote at the beginning of my discussion of the former that Wittgenstein's illustration of IEM invited confusion between the two phenomena. And there has indeed been mistaken conflation of avowal and IEM - beginning I think with Descartes himself! In the Second Meditation, he famously asks 'What then am I?' He expands on the famous answer - the res cogitans is a being that doubts,

¹ Let us say we ask him in Spanish for the meaning of the English sentence 'I am in pain' and he looks blank and uncomprehending.

understands, asserts ... etc. (Anscombe and Geach ed. [1970] p. 70):

These are a good many properties - if only they all belong to me. But how can they fail to? Am I not the very person who is now 'doubting' almost everything; who understands something [etc.]

His emphasis suggests (correctly) that self-ascriptions of these psychological properties are IEM. But later in the same paragraph, after denying that any circumstance could alter this immunity to error, he elides the phenomenon with that of avowal:

... I am now seeing light, hearing a noise, feeling heat. These objects are unreal, for I am asleep; but at least I seem to see, to hear, to be warmed ... (ibid. p. 71).

This is a conflation, because although 'I am now seeing light' might be falsified by the later discovery that I had been dreaming, it could not be mistaken because it was someone else and not I who was seeing light.

But then Descartes was writing before the advent of the exceptionally precise and accurate philosophical tools of the modern age. How surprising then that the conflation continues to be made! Anscombe, at least, is quite innocent on this score. She writes:

The 'I'-thoughts I've been discussing [and alleging to involve a non-referential use of 'I'] have ... not [been those] of Descartes [viz. those expressed in avowals, roughly] ... the propositions that were most difficult for him are most easy for me ... ([1975] p. 63).

That is, Anscombe is clear that the 'I'-thoughts in which she has been interested are not 'incorrigible self-ascriptions of immediate experience' or some such category. But one should not be too generous towards her, since as we have seen on pp. 10-12, she is far from clear just what IEM is! Hacker, though, comments on a possible ambiguity in Wittgenstein's account of 'I' both in the Blue Book (as we have already noted) and in the Investigations; an

ambiguity between a 'no-reference' view of 'I' and a non-assertoric account of avowals. However, just prior to this point in his discussion, he seems to have conflated the allegedly non-referring 'I'-as-subject use, which is correlative with IEM, and the use of 'I' in avowals. He writes:

... it is as a result of the peculiar features of sentences in which 'I' is used as subject - i.e. avowals - that we are led to think that there is a form of empirical knowledge that is non-evidential or immediate ([1972] p. 262).

The mistake here lies in the implication that 'I' is used 'as subject' alone only in avowals, whereas as Hacker himself notes later on, "'I" is used without identification or recognition of a person in a very much wider range of cases than avowals' (ibid. p. 270).¹ That is, IEM is apparent in avowals, but not only there; the class of IEM-utterances is wider than, not merely co-extensive with, the class of avowals.

Holtzmann and Leich make the same mistaken conflation when, in the curious introduction to their [1981], they write:

The 'subjective' use (of 'I') is not based on observation and it simply expresses how things are with the speaker ... It is this use which fails to be referential. But there is another use of 'I' as well, the 'objective' one: this use is based on observation ... ([1981] p. 12).

They go on to say, in a footnote, that 'a similar duality of use is presumably discernible, in Wittgenstein's view, in "joy", "pain", "... means p by S"', where the 'objective' use will involve mastery of 'referential relations', or at least knowledge of the contribution of the respective expressions to the truth-conditions of sentences in which they occur. But these comments arise from a failure to distinguish between judgement not based on identification

¹The mistaken conflation is less apparent in the much-revised second edition (see [1986] p. 281); see also Hamilton ([1987] forthcoming).

(inter alia, IEM judgements), and judgement not based on observation (inter alia, avowals - if they are judgements at all¹). The 'duality of use' of 'I' will then, on some versions of the 'no-reference' account, imply a similar duality in the use of those predicates that figure in both IEM- and non-IEM-judgements. But unless we think that sentence-tokens of the very same type can feature both as avowals and as non-avowals, we will not see how the terms in Holtzmann and Leich's list have dual uses. 'Pain' certainly has no observational use in the 1st-person and it is doubtful that 'joy' does. Of course, there may well be a duality in the use of these predicates between the 1st-person (all 1st-person uses) and the 3rd-person; but this has nothing to do with 'I'-as-object / subject and IEM.

The conflation I have been discussing is of some dialectical significance. In the Philosophische Bemerkungen, Wittgenstein notes how the use of 'I' is 'one of the most misleading representational techniques in our language ... particularly when it is used in representing immediate experience' ([1975] p. 88). Despite the elusive quality of his thought on this subject, it is possible to argue that Wittgenstein, in his later work, came to see that the 'no-reference' view of 'I' is a blunt instrument for resolving the problem of 'immediate experience' and alighted instead on the non-assertoric thesis of avowals. It might be said that there are three problem-areas associated with the 1st-person. They are, in descending order of generality: guaranteed reference, IEM and avowal. They are related but separate, and it is well to keep them so.

The phenomena of IEM and avowal are not identical, then. What is the correct relationship between them? As I suggested in the account of IEM, it seems that the evidential (or perhaps better, the justificatory) basis for an IEM-judgement can be seen as what was termed an avowable-state. This idea is used as the starting-point for a Greek explanation of IEM in Chapter 2 (pp. 107-8 below). But though it looks correct, more needs to be said by way of refinement

¹ If they are not, a simple alteration in terminology is required here.

than is necessary for the purposes of our dialectic. Briefly the difficulty is this. Any non-IEM judgement may be 'based on' awareness of an avowable-state. Testimony, for instance, will cause some auditory or visual state of the subject which at a certain level of description will be avowable. We need to show how, in the case of IEM-judgements, this connection is in some way 'direct' and 'internal'. The directness comes out in the fact that in the IEM case the connection must not be mediated by the experience of someone other than the subject. But this is a necessary condition only; since it is satisfied also in the non-IEM case of Gentle John (p. 11 above). (Though perhaps sufficient too, if we consider only IEM-judgements and non-IEM tokens of the same type, rather than non-IEM-tokens which can never figure also as IEM-tokens). The internal aspect comes out in the fact that one can never (normally) have the experience reported by an IEM-judgement without being in a position to make some germane avowal.

Much work needs to be done to provide a satisfactory characterisation here; and since, as I said, it is not strictly necessary for the purposes of our dialectic, it will be left undone. The notion of IEM itself now awaits further refinement.

1.4) IEM (II)(i) Avowals are IEM

More work is required on the characterisation of IEM, for two reasons. First, it is not readily apparent how the more limited class of avowals, which at first sight are IEM-utterances par excellence, can be accommodated within (IEM₃). Second, it can be argued that demonstratives display IEM as characterised in (IEM₃), and there are dialectical reasons for settling this matter. (If demonstratives display IEM, any argument to the 'no-reference' view based on IEM will have the implausible implication that demonstratives do not refer either - see p. 94.)

To deal with avowals first. It is not possible to apply the test (IEM₃) to avowals. Even if one assumes, contrary to the purely expressive variant of the non-assertoric thesis, that avowals are structured and thus do have the identification and predication components, it still makes no sense to doubt the former component. It is the hallmark of IEM-utterances, as characterised in (IEM₃), that a doubt about whether, say, it was I who fell out of the carriage, means there's no reason for believing that anyone fell out. But the eventuality which this test presupposes can never arise in the case of avowals. There are, unlike with the IEM-utterances we have been considering, no tokens of the type, say, 'I feel cold', where one could doubt whether it is I who feel cold. In the case of avowals, doubt, whether about the identification or predication component or about the complete utterance, is excluded.

Clearly, however, avowals are immune to error through misidentification of the subject. To accommodate this fact, we can make a move so well-tried in previous discussion that it is by now almost second nature! It is the kind of move made in e.g. 'that doubt is impossible does not imply there is infallible knowledge', but here it is in reverse. Instead of formulating IEM in terms of the consequence of a certain kind of doubt, we can do so in terms of the impossibility of a certain kind of justification:

(IEM₅) An assertion is IEM iff it is nonsensical to suppose that one may justify it in the following way:

By putting the case separately for 'Someone is F' and for 'I am that someone', combining the cases and thus obtaining a justification for asserting 'I am F'.

On this basis avowals can be accommodated within the class of IEM-utterances.

(ii) Are demonstrative judgements IEM?

Now to move on to the question of IEM and demonstratives. Of the two issues raised, this is by far the more significant. In bringing out how 'I' and the demonstratives differ in the way in which they display IEM, something not captured in earlier formulations, the motivation for the 'no-reference' view is more sharply delineated. For consider the following example. A French politician (Wittgenstein PI 336) is observing the progress of Baldy's wagonette at the moment at which the misadventure strikes. 'Mon Dieu!' he exclaims, 'What is this! That man has just fallen out of the carriage!' The question of whether these words were uttered in the order in which they were thought need not detain us here. More important is whether 'That man has just fallen out of the carriage!' exhibits IEM. On the formulations we have tried so far, we must say it does. In the terms of the more intuitive defeasibility formulation (IEM₃), a doubt about the truth of the assertion must enjoin the thought that there's no reason to believe that anyone fell out. I cannot cease to believe that it was that man (the object of demonstration) who fell out, whilst still having reason to believe that someone fell out.

But the behaviour of the demonstrative expression 'that man', and that of 'I' in the original example, do differ in this connection. A new, more restricted formulation will serve to bring this out. First I will couch it in terms of defeasibility, which is clearer but unfortunately cannot accommodate avowals (as we have discussed above):

(IEM₆) When based on a certain type of ground, the assertion of 'Fa' is IEM iff the ground is such that:
 The emergence of doubt about the assertion must enjoin the thought that there is no reason to believe that anyone or anything was F.

This amendment of (IEM₃) brings out the fact that in the case of 'I', no misidentification over time is possible. If an utterance 'I am F' is IEM, then this immunity will be preserved for any subsequent transformation 'I was F' made on identical grounds. If any doubt emerges, this doubt must infect the predication component, if the statement is IEM. If I say 'I was F' (or 'I will be F') then, in the IEM cases, I cannot be wrong about who was (or will be) F. This point is of considerable significance. For a range of 1st-person utterances, it seems there is reference to a subject which we cannot fail to identify correctly over time - a persisting subject, statements about which are immune to error through misidentification. This is a feature which, as Strawson would say, 'gives one pause'. It is a remarkable phenomenon and one which arouses the suspicions of those who, like Anscombe, come to espouse Trojanism.

This aspect of IEM has no analogue in the case of demonstratives, it seems. The disanalogy can be demonstrated by posing a dilemma. Consider the assertion 'That man has just fallen out of the carriage', uttered at time t_1 . At time t_2 , either the same assertion can be made, or it cannot. If it cannot, then it will fail the test (IEM₆) for IEM. A small refinement of that test is required to bring out the presupposition involved in this failure:

(IEM₇) When based on a certain type of ground, the assertion of 'Fa' is IEM iff the assertion and the ground are such that:

- (i) the same assertion can be made at a subsequent time;
- (ii) any emergence of doubt about the assertion by that time will enjoin the thought that there's no reason to believe that anyone or anything is F - provided of course, that no independent grounds for believing the latter have emerged.

(I have used the more intuitive defeasibility formulation here; a simple but tedious emendation along the lines of (IEM)⁵ would serve to accommodate avowals.) If the same assertion can be made, then its exhibition of IEM will be empty; and the thought involved, dubiously demonstrative. For it seems that if co-reference, presupposed by but not identical with IEM, is to be guaranteed at all, it must be by means of a device conferring IEM in a transparently tautological way. For instance, by means of an attributive: 'the man we then referred to when we said "that man has just fallen out"'. (Only if such a device is used in the subsequent assertion can it be certain that we are talking of the same person; and only when we are talking of the same person is there the possibility of IEM.) In any case, it is at best questionable that there is reference in the latter case at all, and thus whether a demonstrative thought is 'repeatable' at a later time.

For further illustration, suppose that I come upon someone in a hospital bed whom I take to be the victim of yesterday's accident (viz. Baldy); and it is in fact Baldy. I say 'That man fell out of the carriage yesterday'. I say this because I remember his appearance. Baldy overhears my remark and comments, quite unnecessarily, 'Yes, I did fall out and not very pleasant it was I can tell you'. He says this because he remembers falling out and is no longer subject to the aberrant self-consciousness (if such it was) which, as we know, catapulted him to fame (or notoriety) in the psychiatric world. Now, in the case of both utterances, mine and Baldy's, there is in fact no error in identification. But the grounds of the former do not bestow IEM on the utterance, whereas in the latter they do. That this guarantee is present in the one case and not in the other, is brought out if we consider what happens when each speaker is given, as it happens, spurious grounds for doubting the identity of the victim. If I am told 'No, that man was not the one who fell out' and I believe it, I do not thereby lose my grounds for believing that someone fell out. Whereas in the 1st-person case¹ these grounds are lost - for the utterance is IEM.

¹ For some further discussion of this issue, see pp. 92-94.

1.5) The Generality Constraint

Our final 'key-concept' requiring a neutral interpretation is the Generality Constraint, a feature of language-use which furnishes the ground for a Greek line of argument. The term is Evans', but the idea and the use to which he puts it are anticipated by Strawson in his Individuals. It is by means of the application of one dimension of the Constraint that Strawson hopes to refute the 'two dualisms' of Cartesianism and the no-ownership theory. His 'way with Cartesianism' involves the following claim:

... it is a necessary condition of one's ascribing states of consciousness, experiences, to oneself, in the way one does that one should also ascribe them, or be prepared to ascribe them, to others ... ([1959] p. 99).

'The main point ... is a purely logical one', Strawson writes:

... the idea of a predicate is correlative with that of a range of distinguishable individuals of which the predicate can be significantly, though not necessarily truly, affirmed. (ibid. p. 99n).

Strawson is here drawing attention to one of the complex of abilities which underlies a subject's comprehending self-ascription; that conceptual ability manifested in the predication of a psychological property of members of a class of individuals to which one belongs. In our pursuit of the essential claim we shall next turn to Evans, whose primary concern is with the nature of thought rather than language.

The Generality Constraint is introduced by Evans as the 'one fundamental constraint that must be observed in all our reflections' on the nature of thought. (In these reflections 'we have little enough to go on', even while we admit this Constraint, he writes (op. cit. p. 100).) Evans develops the Generality Constraint from a consideration of what it means to say that thoughts are struc-

tured. He elucidates the case of thought by means of the case of language (which is in fact my primary concern, for reasons that will become apparent). Thoughts are structured in the sense that sentences are structured; understanding a sentence and entertaining a thought each involves a complex of abilities, not a single ability. What complexes of abilities are these? We may set aside for present purposes the abilities associated respectively with performing a particular kind of speech-act and with adopting a particular kind of propositional attitude. It may well be correct to view the attitude towards or the force attached to a propositional content as a further structural dimension of a thought or sentence, one which is subject to the Generality Constraint. Evans himself, in his criticism of Donnellan's analysis of belief, encourages this view (*ibid.* p. 105). But the abilities we are presently concerned with relate to the internal structure of the propositional content. Evans' elucidation proceeds as follows: When we say that a subject's understanding of a sentence 'Fa' is the result of a complex of abilities, we speak (in the first instance) of his understanding of 'a', and his understanding of 'F'; we commit ourselves to predictions about what other sentences the subject will understand, given further postulations about his understanding, and we commit ourselves to a common though incomplete explanation of his understanding of a range of different sentences. So, if he understands 'Fa', and he understands 'G', he will understand 'Ga'; and his understanding of 'Fa' and 'Ga' will be partially explained by his understanding of 'a'. Analogous commitments operate when we say that thoughts are structured, and that a subject's apprehension of the thought that a is F involves a complex of abilities, Evans continues. (I will not detain the reader by spelling out these commitments.) The essential point, he maintains, is that if a subject can entertain the thought that a is F, then, for every property G of which he has a conception, he must be able to entertain the thought that a is G; and likewise for any object b of which he has a conception. But here our slight problem arises. The Generality Constraint is meant to form a basis for an anti-Trojan argument. But if the Constraint is presented as one operating on object-directed thought, it is open for the Trojan to reply at the outset that 'I'-thoughts are not object-directed - so they are ungoverned by any such principle.

Indeed it is hard to see how, when we describe the Generality Constraint as something pertaining to thought rather than to language, we can avoid this kind of ontological commitment. We can, it is true, employ a formulation involving the description of thoughts exclusively in terms of the utterances by which they may be expressed, e.g. 'the thought expressed by "a is F"'. But this will be just an indirect way of talking about language, rather than thought; and we may as well take the direct way. (The feeling that we can never avoid the above sort of ontological commitment in the case of thought, except in the preceding trivial way, derives from the idea that there is no room for a distinction between surface and deep structure in our discussion of thought. In the case of sentences, however, there is a deep structure which surface structure may belie and of which the ontological commitment may be controversial - as in the Homeric Opposition.)

The querulous reader may now interject 'What is all this "dialectical" nonsense! Either the Generality Constraint, as formulated by Evans, obtains, or it does not obtain. Whether or not the constraints on the thought that a is F apply, cannot depend on what is required for the purpose of some so-called Homeric Opposition'. However, my concern is not with the truth or falsity of Evans' claim about the structure of thought, but rather with finding some connected claim that will furnish the basis for a powerful anti-Trojan argument. The formulation Evans proposes may well hold; however, unless some argument is offered for believing that 'I'-thoughts are object-directed, there is no reason to maintain that they are subject to the Constraint. But of course the object-direction of 'I'-thoughts is precisely what is at issue. So we should now turn to a different formulation to that of Evans, one which the Trojan would be hard put to deny, but which will result in some difficulties for him.

This new 'linguistic' formulation of the Generality Constraint proceeds from a slightly diluted version of Strawson's original thought that the idea of a predicate is correlative with that of a range of individuals. It states that for any subject-term and for any predicate, a competent language-user must have the ability to apply a range of predicates to the subject-term, and a range of subject-terms to the predicate. (We need to keep things at the

linguistic level, and refer to subject-terms and not individuals, to avoid begging the question against the Trojan.) The justification for this claim could be the generality implied in the notion of an ability; something which can be exercised on a variety of occasions. Understanding a concept or (in Evans' terminology) an Idea¹ is essentially an ability that can be manifested only sententially, and in a range of sentences employing a range of different constructions. (Whether this requirement is holistic or merely molecularist obviously depends on one's account of these sorts of theory of meaning.)

We shall see in Chapter 2 how an anti-Trojan argument can be developed from the premiss of a Generality Constraint of this sort. This argument will also be separated from another offered by Evans in the same connection.

¹'Idea' stands to 'object' as 'concept' stands to 'property' (see Evans ibid. p. 104).

CHAPTER

TWO

In Chapter 2 I first define Trojanism and relate it to better-known doctrines such as behaviourism and solipsism. I then build on the accounts of key-concepts in Chapter 1 to present arguments for and against the Trojan viewpoint. The Generality Constraint furnishes apparently anti-Trojan material (and is indeed claimed by Evans to do so); but no convincing, straightforward refutation is constructed. However, a line of argument which develops some of the points Evans was making here and which may lead to Trojan difficulties, is indicated. Next, the concepts of guaranteed reference and IEM furnish the basis for arguments in favour of Trojanism, the former via consideration of the situation of a victim of Anscombe's sensory-deprivation tank. (That the former does so despite our analysis of the phenomenon in Chapter 1 as 'empty' seems to indicate that there is some unresolved tension at this point.) After elimination of mistaken presentations and some often painstaking refinement, arguments emerge which are plausible though far from conclusive. Doubt is particularly cast on the assumptions behind the first argument (Anscombe's Tank Argument). The chapter concludes with the claim that 'no-reference' Trojanism cannot be assessed without assessing the more important Trojan claims about avowals, to be discussed in Chapter 4.

2.1) Historical Account and Definition of 'No-Reference' View - Wittgenstein's 'Philosophical Remarks'

(i) What is Trojanism?

The loci classici of the 'no-reference' view are to be found in the middle period works of Wittgenstein - but although the view surfaces with greatest explicitness in these works, Trojanism is a continuing thread from the Notebooks 1914-16 to the Philosophical Investigations. Clearly, the espousal of Trojanism by students of Wittgenstein such as Anscombe and Malcolm is no historical accident.

It is in four works of or pertaining to the period 1929-33 that Wittgenstein's Trojanism appears at its starkest - the Philosophical Remarks, Wittgenstein und der Wiener Kreis, 'Wittgenstein's Lectures in 1930-33' by G.E. Moore, and Wittgenstein's Lectures, Cambridge 1932-5 ed. A. Ambrose. In these accounts, in contrast to the compressed and recondite presentation of the Tractatus, Trojanism moves into sharp and detailed focus against the background of the verificationist semantics of the Logical Positivist phase. In the Blue Book and Notes for Lectures of 1933-6, as the Tractarian and verificationist elements are on the wane, there is a movement towards the less explicit Trojanism of the later period. In his [1972], P.M.S. Hacker characterises the 'metaphysics of experience' of the 1929-33 period as 'methodological solipsism', in contrast to the 'transcendental solipsism' of the Tractatus. From 1932-33, methodological solipsism is, he claims, abandoned in favour of the metaphysics of experience which came to be expressed in the Private Language Argument.¹ The question of how far the later philosophy is still essentially solipsistic is a piquant one however; it is addressed, and answered affirmatively, by David Bell in his unpublished paper 'Solipsism'. It would be less contentious to say that the later philosophy is still Trojan; the

¹ The retractions in Hacker's revised edition [1986] do not essentially affect these claims.

nature of the relation between Trojanism and solipsism is explored in the next section. It is the aim in the present section to give an illustration, and as far as possible a definition of Trojanism; but no definition of Trojanism nor of solipsism quite allays the fear of the present writer that it is not possible really to know, of either of these strange doctrines (or its negation), what it means. In this respect, the Homeric Opposition conforms to the most refined canons of philosophical dialectic.

Rallying against this negative mood, let us ask 'What then is Trojanism?' Clearly, we can say as a start that it must involve some form of opposition to the 'straightforward construal' of a range of 1st-person utterances (viz. IEM-utterances or the avowal sub-class of these). It sees these as functioning in a quite different way from the way 3rd-person utterances (and other 1st-person utterances) function - an asymmetry denied by the 'straightforward construal' (which is the 'Greek' view). To characterise Trojanism in a less imprecise way, it is necessary to look at its historical role as, if you like, the 'obverse' of Cartesianism. Or, if that expression takes its affinity too far, then as a 'reaction' to Cartesianism. Trojanism takes the threat of a Cartesian viewpoint sufficiently seriously as to warrant a very radical response - which Greeks find unnecessary and in fact counter-productive. Quite where the Cartesian threat is located, and the kind of radical response required, depends on the variety of Trojanism. The middle-period Wittgenstein located the danger in a referential account of 'I' for IEM-utterances, which he therefore denied. (An historical parallel to this, but one to be handled with care, would be Lichtenberg's reaction to the cogito of Descartes himself.) The later Wittgenstein came to place the emphasis more on the status of avowals, rejecting their 'straightforward construal' as assertoric (or rather, their construal as straightforwardly assertoric - see Chapter 4).

It would be wrong, however, to be too dogmatic about the change in Wittgenstein's Trojanism - simply because of the (for the most part) highly suggestive nature of his treatment of the issue. Hacker writes that 'it is noteworthy that [the] distinction between the use of 'I' as subject and as object [in the Blue Book] does not reappear in the later writings' ([1986] p. 223n). But what we have

is really a change of emphasis rather than a rejection of 'no-reference' Trojanism in favour of Trojanism about avowals. It is possible, however, in the light of the undoubted changes in Wittgenstein's philosophy of language, to characterise changes in his position on the 'no-reference' view per se. The key notion which I will select in order to achieve this purpose is that of logical form. Crudely, before 1929 Wittgenstein was a realist about logical form. After 1933, he was not. 'Realism about logical form' I take to be the view that any sentence-type possesses a semantic deep structure which mirrors the articulation of the facts - whether these are the facts which constitute the world, which is the Tractarian view, or the facts about understanding located in the mind or brain of the language-user, which is the view a latter-day realist such as Davidson seems to be compelled to accept.¹ In his pre-1929 Trojanism, therefore, and to a confused extent in that of 1929-33, Wittgenstein is concerned to counter the Greek account of the logical form of 1st-person sentences with an alternative Trojan account which accepts the realist terms of that of the Greek. The debate here takes the form: 'Must the logical form of such sentences, in mirroring the facts, comprise a grammatical subject that is also an object, or not?' After 1933, the debate will take the form: 'Is there a logical form at all which mirrors the facts and thereby comprises a subject which is also an object?' Wittgenstein's position in the former debate is clear even if one regards the terms of that debate as quite unsatisfactory. That Wittgenstein's answer to the later question is 'no', is not incompatible with his continued espousal of the 'no-reference' view (which however becomes increasingly a relic of an anti-Cartesian response which the non-assertoric thesis of avowals renders outmoded - see p. 141). It is true that, realism over logical form superseded,

¹ Davidson's programme involves investigation of the logical form a sentence-type must have if the speaker's understanding of it is to be satisfactorily explained. The laconic comment on what Davidson is compelled to accept is controversial (but peripheral to my concerns); the philosophical world has apparently failed to be struck by how bizarre and eccentric an enterprise the Davidsonic programme really is.

the point at issue in the first stage of the Homeric Opposition becomes less intuitive. But the case discussed under IEM, for instance, does not require realism over logical form - nor do any Trojan cases based on the 'epistemology of reference', as one might term it.

It may be felt that some of the ideas in the preceding paragraph - notably that of 'logical form' - are elusive. The unease here is the same as that felt when the question 'What is reference, anyway?' is being asked. The intuitive grasp of the Homeric Opposition seems under threat, we are no longer clear what is at issue, nor indeed that there is any real issue. It would be preferable, at this point, rather than retreating to backgammon, to read on - and if the talk of logical form remains unilluminating, it may turn out not to be essential.

(ii) The 'Philosophical Remarks' account

The most extended first-hand statement of Wittgenstein's middle-period Trojanism occurs in the Philosophical Remarks, paras. 57-58. I will outline this discussion at length, since many of the characteristic aspects of Wittgenstein's Trojanism are present.

The concentration on avowals among the range of 1st-person IEM-utterances, demonstrating, as discussed elsewhere, that equivocation in Wittgenstein's accounts between a Trojanism applying to avowals, and one applying to all IEM-utterances, is apparent from the outset:

One of the most misleading representational techniques in our language is the use of the word 'I', particularly when it is used in representing immediate experience, as in 'I can see a red patch' (ibid. p. 88).

Wittgenstein goes on to suggest that if, in thought-experiment, we abandon this 'technique' in favour of one in which the 1st-personal pronoun is dispensed with, we will see that the earlier representation 'wasn't essential to the facts':

Not that the [new] representation would be in any sense more correct than the old one, but it would serve to show clearly what was logically essential in the representation (ibid.).

This new notation might be termed the 'oriental-despot language game', since Wittgenstein brings out the 1st / 3rd person asymmetry which Trojanism takes so seriously by asking us to imagine a despotic oriental state in which the despot is linguistically privileged, in the following way. When, as we would say, the despot (the 'centre' of the language-game) is in pain, he says 'There is pain'. When he observes that, as we say, one of the subjects is in pain, he says 'X is behaving as the centre behaves when there is pain'. Wittgenstein writes:

It's evident that this way of speaking is equivalent to ours when it comes to questions of intelligibility and freedom from ambiguity ... (ibid.)

That is, 'I' can be eliminated from the language and a substitute way of talking supplied, without loss of the language's expressive power - or so Wittgenstein seems to assume. As we shall see, doubt is cast on this assumption by the charge that 3rd-person ascription is now behaviouristic - pp. 39-40. More seriously, the further we move from the sensation-paradigm to mental states such as thinking, the less straightforward it becomes to envisage the kind of behaviour that might be involved. It should be noted that 'the centre' refers to a body; the despot's 3rd-person ascriptions contain an indirect reference to the behaviour of this body, and his 1st-person non-psychological ascriptions refer to it directly.

It is not clear that the image of the despotic state (alluded to both at PR 58 and in Waismann [1979] p. 49) is helpful however. For as Wittgenstein says, 'this language could have anyone at all as its centre'. We could therefore envisage a situation in which each is the centre of his own language, is his own 'despot', (pace Bell ibid. p. 12); if that is the correct way to put it. What will interest us, then, is the comparison (i) between our language and an arbitrarily chosen member of this set of PR languages, and

(ii) between individual members of the set.

With regard to the latter, Wittgenstein comments that the language with myself as the centre has a 'privileged status', but this is not something that can be described:

The privileged status lies in the application, and if I describe the application, the privileged status again doesn't find expression, since the description depends on the language in which it is couched (ibid.).

Clearly, the Tractarian distinction between saying and showing is in operation here. The idea seems to be that since, for any language with himself as centre which a speaker might adopt, its 'particular adequacy' for himself is presupposed by the very form of discourse used, it cannot be expressed in that language. And, Wittgenstein notes, 'in the terms of another language my language occupies no privileged status whatever'. It is not clear what these claims amount to, however. We can certainly say the kinds of things about the individual PR-languages that Wittgenstein has said in comparing our present egalitarian discourse with his inegalitarian proposal. Isn't it the case that the language with myself as centre is particularly adequate for me because only in that language is the non-referential nature of self-conscious self-ascription of IEM-states (as we say) reflected at all? If I myself adopt the language which has Jones as centre, I may well, as Bell maintains (ibid. p. 12) be able to say, when in pain, 'I am behaving as Jones / the centre behaves when there is pain', but this locution hardly serves to bring out the epistemic facts of the matter - as the inegalitarian way of speaking is meant to.

What of the former comparison, between our language and an arbitrary PR language? Wittgenstein's claim is that the two notations are 'equivalent'; anything that can be said in one can be said in the other. The possibility that this might not be so is not examined. But this is just the question. The thought-experiment, presupposing as it does that the inegalitarian language is 'fact-preserving' (in Bell's phrase, ibid. pp. 6ff.), is merely an illustration of what Trojanism might involve, were it correct, and not an argument for that viewpoint. (Though if one found the picture

coherent, that fact might be persuasive.)

(iii) The danger of 'covert behaviourism'

We need now to return to the question of whether Wittgenstein's inegalitarian proposal is behaviouristic. In answering this question we may advert initially to that worthy piece of journalism, 'Wittgenstein's Lectures in 1930-33' by G.E. Moore. Moore's despatches as Foreign Correspondent for Mind in the Viennese thought-world are painstakingly assembled, and deal explicitly though laconically with the issue of behaviourism. He writes:

He seemed quite definitely to reject the behaviourist view that "he has toothache" means only that "he" is behaving in a particular manner ... Later on, he said that we conclude that another person has toothache from his behaviour, and that it is legitimate to conclude this on the analogy of the resemblance of his behaviour to the way in which we (I) behave when we (I) have toothache ([1955] p. 12).

It is clear, also, from the reference to the 'controversy over behaviourism' in the PR itself (p. 94) that at this time, as well as more famously later on, Wittgenstein is concerned to escape the charge that he is a 'covert behaviourist'.

Whether he can escape it is another matter. If it is behaviourism he is committed to, then clearly it is a kind of behaviourism that avoids the problems with the 1st-person that make conventional behaviourism so implausible. Bell (ibid. p. 19) provides a defence of the view that it is not behaviourism at all; but it looks like he is batting on a sticky wicket. For let us suppose, as Wittgenstein intends, that our language and any arbitrary PR language have the same factual capacity. (Any fact that can be reported in one can be reported in another.) Let us suppose also (as seems natural) that the relation 'having the same factual capacity as' is transitive. Then any two PR languages, with different 'centres', will have the same factual capacity as each other. Which must entail that:

'There is pain'

said in L_1 makes the same statement as

'X is behaving as the centre behaves when there is pain'

said in L_2 of the subject which is in fact 'the centre' in L_1 . QED? On the face of it, yes - but one can't help feeling that some further subtleties beyond the scope of the present discussion are called for in dealing with the vexed issue of behaviourism. Nonetheless, as things stand, a dilemma now opens for the Trojan which takes us neatly on to the next part of our discussion. It looks, in brief, as if the Trojan must endorse either behaviourism or solipsism. Say he wishes to maintain 'fact-preservation' between our language and the PR languages. Then he will have to deny the principle that if the utterances (the two quoted above, say) are to make the same statement, their constituent expressions must co-refer. For 'There is ...' and 'X ...' manifestly must not co-refer (on the Trojan account); yet they are to make the same statement. If the Trojan dislikes the behaviourist conclusion we have drawn, he will have to endorse the 'same statement, same reference' principle, forget about 'fact-preservation' and endorse solipsism. The argument spelling this out will be presented on pp. 44-45 below, after the varieties of solipsism have been discussed.

2.2) Solipsism and the 'No-Reference' View

(i) Epistemological and metaphysical solipsism

Nothing has so far been said of the extent to which Trojanism might be solipsistic. What, then, is the relation between the 'no-reference' view and solipsism? First, we need to know what are the varieties of solipsism. A useful distinction at the outset, one stated explicitly by Nagel [1970] and Bell (unpublished article), is that between epistemological and metaphysical solipsism. (Pears ([1975] p. 273) notes this distinction,¹ and Hacker ([1972] p. 186) seems also to be aware of it.) In each version, the range of truths said to be knowable or expressions said to be intelligible, can vary with the degree of 'liberalisation' of the theory. Pears claims that one limit is that 'the privileged basis cannot include any things which would not be accessible in any direct way' (*ibid.* p. 273). The opposite illiberal extreme is a 'solipsism of the present moment'.

I mention the latter dimension merely to set it aside, at least temporarily. For what we want is 'essential solipsism' - which I take to be solipsism in its most attractive form; possibly 'the truth in solipsism', but at least 'the plausibility in solipsism'. This I assume to be related to one side in the Homeric Opposition.

For this purpose we need to invoke the more important initial distinction. For the solipsism which most authorities scorn or scoff at is epistemological solipsism. It is this of which Schopenhauer memorably writes:

It can never, of course, be demonstrably refuted. Yet in philosophy it has never been used positively other than as a

¹ He refers to 'linguistic', rather than 'metaphysical' solipsism.

sceptical sophism, i.e. as a pretence. As a serious conviction, on the other hand, it could only be found in the madhouse, and as such needs a cure not a refutation ... (quoted in Bell ibid. p. 2).¹

This is the 'traditional attitude' to solipsism of which Bell writes; it is endorsed by F.S.C. Schiller, among others. ('The non-existence of solipsists by no means seems to me devoid of ... significance ...' he sniffs.) But as Bell notes, this attitude had to be abandoned by those exposed to Wittgenstein's ideas in the 1920s and '30s - 'in face of a growing realisation that there had been, perhaps even still was, a solipsist'. If this were a genuine realisation, though, the solipsist in question was a metaphysical solipsist.²

What are these two kinds of solipsism? We can say this. 'Epistemological solipsism' makes some claim about the grounds for the application of the sortal concept 'person'; a concept which can, with significance though not correctness, the epistemological solipsist will maintain, be applied by oneself to oneself and others. However, he would have us believe that, for each of us, there is insufficient evidence to warrant the belief that there are instances of 'person' other than oneself. E.g. 'pain', 'intelligence' cannot with justification be applied to those I took to be persons like myself. So this is the form of the classical dispute over epistemological solipsism:

¹ An interesting discussion of Schopenhauer on the self is to be found in C. Janaway [1984].

² At least one other possible solipsist has suffered philosophical neglect, however. Charles Mingus, the great jazz bass-player, composer and band-leader has gone on record as saying:

I was born an extension of life itself - the whole outdoor scene, sky, moon, sun, universe, space, the whole scene is me ... Charles Mingus. (Jazz Calendar 1985)

Whether this is an affirmation of a philosophical view rather than a mere expression of the well-attested Mingus egomania is a moot point, however.

'Solipsist': 'The concept 'person', and its cognate, harmless and salutary companion 'self', is unified in sense, but disintegrates in application'.

'Realist' : 'No - it is unified in both ways. The grounds for applying 'person' etc., are good enough, because we can (shakily) 'infer' the existence of other persons, or, better, assert the existence of a primitive concept 'person' such that it cannot fail to apply to others'.

Epistemological solipsism is, as Bell suggests, a form of scepticism; its propounder is sceptical about the unity of application of 'person'. That is, the difference in quality of 1st- and 3rd-person grounds for the application of 'person', should lead one to be sceptical that there is a justified 3rd-person application.

To this, there are two responses, at least. First, the classical Strawsonian response given by the realist. But second, a more profoundly sceptical response: scepticism over whether the concept 'person', etc. is 'sound'. Is it not rather that there is a duality in sense, and hence, if one likes, in application? It is this contention that is held in common by metaphysical solipsism_S and metaphysical solipsism_P (MSS and MSP). And it is plausible to hold that essential solipsism consists in just this view of the duality of sense between 1st-person (or rather, always, some 1st-person) utterances and 3rd-person utterances. It is very rough, but the idea is: 'I am in pain' and 'X is in pain', said of me, do not make the same statement. This, on the grounds that the subject-terms are not co-referential (MSS); or that the predicate-terms are not (MSP).

What is now required is a better formulation of metaphysical solipsism, and of its relation to epistemological solipsism, and of the nature of its scepticism:

MSP : No unified concept 'person' can be obtained such that the doubts or beliefs of the epistemological solipsist can be framed, on the ground that there is no range of predicates distinctive of the concept and applicable to any of its alleged instances.

Object x and y fall under a common sortal only if there is a range of predicates applicable to x and y which is distinctive of that sortal. With Strawson, we took it that there is a range of predicates which can typically be applied to persons - viz. P-predicates ('P' for 'Person', not 'Psychological'). This set includes any predicate applicable to a person but not to a material body; this will extend beyond the 'psychological', to e.g. posture, motion etc. ([1959] p. 104). But it turns out that there is no such range of predicates - rather, the sense of these expressions differs depending on whether they figure in 1st- or 3rd-person utterances. The nature of the different senses will obviously depend on the account we give of the epistemic asymmetry which generates them.

MSS arrives at the same position via a different route:

MSS : No unified concept 'person' can be obtained such that the doubts or beliefs of the epistemological solipsist can be framed, on the grounds that 'I', in typical cases of purported psychological ascription at least, does not refer, a fortiori¹ to a person.

The first clause, common to MSP and MSS, when suitably explicated in terms of 'duality of sense', gives us what I have termed 'essential solipsism'. It has been well said that metaphysical solipsism 'transforms an epistemic asymmetry within a realist framework, into a semantic asymmetry within a verificationist framework' (C. Wright). The senses of 1st- and 3rd-person ascriptions are now distinct and given by the distinct methods of verification of each.

(ii) The connection between Trojanism and solipsism

We are now in a position to state a simple argument connecting solipsism with Trojanism. It is this. Let us assume (as seems uncontentious) that if two utterances make the same statement, their

¹ This amounts to, or entails, Strawson's 'No-ownership' view of psychological states.

constituent expressions must co-refer. Now, consider again 'I am in pain' and 'X is in pain' said of me. If these two utterance-tokens make the same statement, any reference involved in one must be involved in the other. Ex hypothesi, this is not so. So I alone can express the content of 'I am in pain'. Others have no way of saying, let alone knowing, what I express when I say 'I am in pain'. To say this is simply to claim that 'no-reference' implies 'dual sense', which itself implies or constitutes solipsism, as was previously indicated. Whether this conclusion is welcome to the Trojan (or rather, to which kind of Trojan it would be welcome), and what options he has if it isn't, is discussed below (pp. 50-51). (Though I continue now with a few psychological remarks about the connection.)

The kind of solipsism Trojanism is, arguably, connected with is of course metaphysical solipsism. Many Trojans, including the most famous living one (Professor Anscombe), would probably want to resist the idea that they were solipsists at all; but that might be due to the justifiably poor reputation of epistemological solipsism and to the fact that metaphysical solipsism is often not regarded as solipsism at all. Epistemological solipsism is indeed a consequence of Cartesianism and its account of the 'privacy of experience'; so it is plausible to see in the varieties of solipsism an analogue of the dialectic discerned in the first section of the present chapter (p. 34). That is, just as Trojanism is a response or reaction to Cartesianism, so is metaphysical solipsism to epistemological solipsism. Given this, and our preceding account, it cannot be correct to suspect as Nagel does, that 'the epistemological version of solipsism ... is ultimately dependent on the metaphysical version' ([1970] p. 104n).

It is well to be cautious in our generalisations in an area notorious for its production of philosophical nonsense, however. Although metaphysical solipsism may be a reaction to epistemological solipsism, it is far from clear that it is a reaction of exactly the same kind as that of Trojanism to Cartesianism. Wittgenstein's discussion, recorded by Ambrose in her [1979], brings out some of the complexities in the matter.

(iii) Wittgenstein on 'what the solipsist wants'

The distinction Wittgenstein is making looks very much like that already drawn between epistemological and metaphysical solipsism. He havers over whether the latter is solipsism at all - as we shall see. But this is not a crucial matter, in fact. He says:

The solipsist who says 'Only my experiences are real' is saying that it is inconceivable that experiences other than his own are real. This is absurd if taken to be a statement of fact (ibid. p. 22).

Likewise with the person we might term the realist:

'Only my experiences are real' and 'Everyone's experiences are real' are equally nonsensical (ibid. p. 23).

This is a dialectic which occurs also in PI 402 - which we will come to. The sceptic and his opponent conduct a mistaken debate at the factual level. Wittgenstein wishes to undercut this by showing what the solipsist ought to be saying - presenting his statement as 'a grammatical proposition', 'the statement of a rule':

The solipsist wishes to say, 'I should like to put, instead of the notation "I have real toothache", "There is toothache"'. What the solipsist wants is not a notation in which the ego has a monopoly, but one in which the ego vanishes (ibid. p. 22).

This is because in saying 'Only I have real toothache', "'I'" is no longer opposed to anything' (ibid. p. 22). The view in which the ego vanishes is Trojanism, of course. And with the talk of the 'grammatical proposition', and shortly afterwards of 1st- and 3rd-person propositions being on a 'different level', it looks as if Wittgenstein's view also involves that transformation of an epistemic into a semantic asymmetry which is definitive of metaphysical solipsism.

The factual / grammatical dichotomy of interpretation is of course a favourite one. A close analogy is with PI 247 ('Only you

can know if you had that intention'). The propositions in question are factually incorrect, or absurd - yet they can be used in explaining the meaning of 'toothache', or 'intention'. This is paradoxical; but in the 'intention' case, Wittgenstein locates the paradox by means of a gloss on 'know' ('And here "know" means that the expression of uncertainty is senseless'). In the latter, the paradox serves to bring out the incorrigibility of self-ascription of intention (in the standard cases at least); of course others can be said to know what my intention was, but not in the incorrigible way that I can. But what does 'Only I have real toothache' treated as a 'grammatical proposition', serve to bring out? Perhaps, as Wittgenstein later says, that 'I have toothache' and 'He has toothache' are 'on a different level' (ibid. p. 23). There is a later treatment of the issue in the Investigations :

... this is what disputes between Idealists, Solipsists and Realists look like. The one party attack the normal form of expression as if they were attacking a statement (Idealists, Solipsists); the other defend it as if they were stating facts recognised by every reasonable human being (PI 402).

The protagonists here lock horns in a futile engagement at the factual level - as Bell comments, 'this is what the debate between sceptical solipsists and those with a Russellian "robust sense of reality" looks like' (ibid. p. 17). Set against this description is the 'metaphysical solipsist' account of para. 403:

If I were to reserve the word 'pain' solely for what I had hitherto called 'my pain', and others 'L.W.'s pain', I should do other people no injustice, so long as a notation were provided in which the loss of the word 'pain' in other connections were somehow supplied. Other people would still be pitied, treated by doctors and so on ...

But what should I gain from this new kind of account? Nothing. But after all neither does the solipsist want any practical advantage when he advances his view!

Here, as in the Cambridge Lectures, it is unclear whether Wittgenstein thinks his proposal is solipsistic. His 'havering' over this question in these earlier Lectures is brought out when he says

'Getting into the solipsistic mood means not using the word "I" in describing a personal experience':

Acceptance of such a change is tempting because the description of a sensation does not contain a reference to either a person or a sense organ (ibid. p. 22).

Tempting it may be, but he later seems to retract his point about 'the solipsistic mood' (a nice expression!) when he says that 'the solipsist does not go through with a notation from which "I" or "real" is deleted' (ibid. p. 23). The question really is one of 'semantics' in the pejorative sense, however - the material issue is clear enough.

From the preceding discussion it should be apparent that there are significant connections between, on the one hand, Cartesianism and epistemological solipsism, and on the other, Trojanism and metaphysical solipsism. Further, one may conclude that there is one variety of Trojanism which is metaphysically solipsistic - what variety this is will be discussed in the next section in the context of that argument against Trojanism derived from the Generality Constraint.

2.3) Arguments against Trojanism: from the Generality Constraint et al.

(i) Militant and moderate Trojanism and the 'continuous ability' challenge

Evans in his [1982] believes that the Generality Constraint implies a rejection of Trojanism (as we shall see). But what he gives as an argument turns out not to have anything essentially to do with the Constraint. However there is a different argument which does. We recall that the Constraint states that the subject must, inter alia, be able to apply a range of subject-terms to any given predicate - say, 'feels cheerful'. Now, when the subject applies a 3rd-person pronoun to this predicate, and says of someone 'He is cheerful', the ability he manifests is clearly referential; he has the ability to use a referring expression. Indeed, his ability in using a range of subject-terms is referential - except, if the Trojan is correct, when it comes to the 1st-person pronouns. So the challenge to the Trojan is this: What is the explanation of this requirement, extending as it does to the 1st-person, if the ability in question is discontinuous between 1st- and 3rd-person contexts (i.e. is of a different character)?

This challenge has a similar form to another, often mistakenly thought to constitute a refutation of Trojanism, and by which Anscombe has been decidedly untroubled (see Appendix: Elizabeth Anscombe und der Bangor Kreis, p. 240). The challenge in this latter case is the following: In 1st- and 3rd-person sentences, there are occasions on which 'I' on the one hand, and the subject's proper name or some description true of the subject on the other, are intersubstitutable salva veritate (provided appropriate changes are made to verb inflections etc.). What is the explanation of such truth-value links between 1st- and 3rd-person sentences, if not the co-reference of the terms in question?¹

¹ I owe appreciation of this point to discussion with Ian Browne.

Now, it is important to bear in mind that neither the argument from a Generality Constraint, nor that from the existence of truth-value links, constitutes by itself a refutation of Trojanism; each is, rather, a request for a story from the Trojan - and we may judge that his story-telling abilities are wanting. There is this disanalogy however: in the latter case the truth-value links and not the co-reference are the data, but in the former it is less clear what exactly the data are. This makes the anti-Trojan challenge weaker there, at first sight; for we are less sure exactly how far the requirement extends.

Now, it was stated in the immediately preceding account of solipsism and the 'no-reference' view, that there is one variety of Trojanism that is metaphysically solipsistic. We shall now see how the different varieties of Trojanism imply different responses to the challenge from the Generality Constraint. First, the militant Trojan - the metaphysical solipsist. He maintains that the predicates in self- and other-ascriptions differ in meaning according to whether the subject-term is referring or non-referring (and thus a 'pseudo'-subject-term). Thus 'pain' in 'I am in pain' and in 'He is in pain' is not univocal; each kind of occurrence can and must be understood in isolation from its counterpart. In contrast, the moderate Trojan maintains that such concepts are univocal, even though 'I', as it figures in IEM and non-IEM ascriptions, is ambiguous, and that 3rd-person transforms of IEM 1st-person ascriptions do in fact make the same statement. He is thus committed to dissolving the connection between identity of reference and identity of statement. If he did not do this he would be committed to solipsism (see the 'simple argument' on pp. 44-45 above). If militant Trojanism is an expression of solipsism, then moderate Trojanism may perhaps be seen as an expression of Nagel's idea that there are subjective and objective points of view, neither of which is reducible to the other (see his [1970]). The distinction between the types of Trojanism in terms of whether or not certain concepts are univocal is over-simple, in fact, because there are ranges of expressions between the straightforwardly univocal and the straightforwardly ambiguous (between the cases of, say, 'scarlet', and 'bank' with its clear two senses). A more subtle presentation of the matter in the case of Trojanism would require comparisons between particular examples. In particular it

may be sufficient for the moderate Trojan to claim that the uses of e.g. 'pain' are just importantly analogous.

The militant and moderate Trojan will respond differently to the Greek challenge. The first will say that there is no Generality Constraint extending across 1st- and 3rd-person ascriptions. He will, however, put forward instead a highly diluted Constraint; and in so doing make a counter-attacking move in weather conditions which, as always, are favourable to the Trojan side. What I have in mind is this. 'I'-thoughts are, like thoughts about the weather, the militant will say, not subject to a Generality Constraint of the sort hitherto discussed - because the subject-term in e.g. 'I am unhappy' is merely apparently such, and is in fact akin to the first term in 'It is raining'. Both are alike non-referential. A speaker's ability to use these expressions is manifested, and his understanding of them tested, by means of variation in tense and predicate alone - there is no dimension of generality akin to that of variation in subject-term for the genuinely referring expressions. So, to say, as Strawson would, that the predicates involved in self-ascription cannot be genuine unless affirmable of a range of subjects is as silly, the militant Trojan concludes, as saying that 'is raining' cannot be a genuine predicate unless affirmable of a range of subjects besides 'it'.

Thus much for the militant Trojan. (The weather, of course, is the Trojan's greatest ally. But also, almost his only ally. The Greek riposte to the above sally would be that whilst the generality claim about the weather is silly, that about 'I' is not - because of the truth-value links, in part, at least.) The moderate Trojan can also, if he wishes, avail himself merely of the highly diluted Constraint. But it would be more consistent of him to allow the extent of the Constraint across 1st- and 3rd-person ascriptions, but deny that it requires explanation in terms of a continuous type of ability. And here we see really that the Greek challenge evaporates. For why should the Trojan not be allowed to say this? It might be claimed, for instance, that a subject needs a grasp of 'Someone is in pain' as well as 'I am in pain' and 'X is in pain', before he can be credited with an understanding of 'pain'. Yet the ability expressed in the use of the quantifier would clearly not be referential. The grasp of the 1st-person ascription might be more

basic than that of the quantified expression, it might perhaps be argued, yet surely it would be somewhat arbitrary to rule that the Generality Constraint did not extend to the latter. Whether or not this particular example is considered, the point remains that the requirement of a continuous type of ability is not a compelling one. The explanation of the conformity of 1st- and 3rd-person ascriptions alike to the Generality Constraint is simply that the subject must be able to apply the predicate to the whole range of appropriate subject-terms. That different kinds of ability are required in particular cases seems neither here nor there.

(ii) Evans' argument: 1st-person undecidable propositions

So much for the principal anti-Trojan argument derived from the Generality Constraint. Now for an argument connected in Evans' mind with the Generality Constraint. The former argument, or challenge rather, asked 'What is the explanation of the conformity of 1st-person ascriptions to the Generality Constraint, on the Trojan account?' Evans' argument seems to ask 'How can the Trojan allow that certain 1st-person ascriptions (undecidable ones) conform to it at all?' Evans' thoughts here are expressed in a very compressed way, and implied in his argument are claims about the 'fundamental level of thought' about persons which are central to The Varieties of Reference.¹ Without going too far into the latter (which would not be a feasible enterprise in the present context) it is possible to indicate Evans' general line of thought and the confusions it involves.

To begin at the end of his discussion. Evans states there that there is an opposition between the Generality Constraint and an 'Idealist conception of the self' - a conception which if it were

¹ He says that one's Idea of oneself involves a 'fundamental identification' of oneself as a person located on a 'spatio-temporal map of the world'. Thus, he claims, 'I'- and 'here'-thoughts, self- and demonstrative identification, are inter-dependent (ibid. p. 211). This the Trojan contests.

proposed 'would be the same as saying that "I" does not refer to anything' (ibid. pp. 212 and 212n). He has earlier indicted various accounts on the charge of an Idealist violation of the Generality Constraint. He began in fact with the question 'What are our "I"-Ideas composed of?', and addressed accounts which are exhausted by the ingredients of action- and information-components. Such accounts leave something missing:

[But] so long as we focus upon judgements which a person might make about himself on the basis of the relevant ways of gaining knowledge, the inadequacy may not strike us. A subject's knowledge of what it is for the thought 'I am in pain' to be true may appear to be exhausted by his capacity to decide, simply upon the basis of how he feels, whether or not it is true (ibid. p. 208).

But not all entertainable 'I'-thoughts have a basis in ways of gaining self-knowledge:

... our view of ourselves is not Idealistic: we are perfectly capable of grasping propositions about ourselves which we are quite incapable of deciding, or even offering grounds for (ibid. p. 208).

Evans then gives examples of 'I'-thoughts which are not directly decidable by the subject, but which, in order to be verified at all, require 3rd-person verification: 'I was breast-fed', 'I was unhappy on my 1st birthday', 'I tossed and turned in my sleep last night', 'I shall be dragged unconscious through the streets of Chicago', [de jure] 'I shall die'. These are thoughts one can grasp; 'in other words', Evans concludes enigmatically, 'our thinking about ourselves conforms to the Generality Constraint'. (He goes on to say in a footnote that this thought is diametrically opposed to Wittgenstein's expressive account of 1st-person psychological utterances.) This conformity requires that the subject know the truth-conditions of $\ulcorner I = \delta_f \urcorner$ where δ_f is a fundamental identification of a person, viz. an identification of a kind which could be available to someone other than oneself. Evans concludes by relating the preceding to the thought at the heart of his Generality Constraint:

Only if [one's 'I'-Idea includes knowledge of what it would be for an identity of this kind to be true] can one's general understanding of what it is for a person to satisfy the predicates ' ξ is dead', ' ξ is breast-fed' ... etc. be coupled with one's Idea of oneself to yield an understanding of what it would be for oneself to satisfy these predicates (ibid. p. 209).

(iii) Trojan response and deficiencies

So Evans is here linking the ability to grasp the 1st-person undecidable propositions with the ability to fundamentally identify oneself as a person among others. The notion of 'fundamental identification' is obviously a technical one, but a fairly rudimentary grasp of the point will suffice for present purposes. Evans goes on to discuss accounts which deny that I can make any such identification - principally Nagel's in his [1970]. This is a fascinating topic - but one cannot discuss everything! The relevance of these issues here is to the question of anti-Trojan arguments based on the Generality Constraint as defined in Chapter 1. Now, at first sight there is no reason why the moderate Trojan, at least, should infringe any Generality Constraint through a denial that one can grasp the undecidable propositions¹ - since he need not make any such denial. It may indeed be the case that Idealism, construed as a doctrine which denies sense to propositions which are undecidable, implies that 'I' does not refer. But Trojanism need not imply Idealism; and though the Trojan owes us an account of the undecidable propositions Evans adduces, there is no reason to suppose he denies that one can grasp them. Perhaps Evans is making a mistake here in his characterisation of Idealism; certainly that well-known species of Idealism (if such it is) known as Anti-Realism does not deny sense to the undecidable, but simply claims that whatever gives meaning to undecidable propositions, it cannot be their ungraspable truth-conditions. It seems thus far that Evans has picked an unworthy opponent.

¹'Undecidable' here = '1st-person undecidable'.

Are the undecidable propositions problematic for the Trojan, then, even if he does not deny they can be grasped? What account will he offer of them? He will say that the significant point is that the undecidable propositions adduced by Evans are all non-IEM; so they will involve the referring use of 'I', applicable to a range of predicates from which 'pain', 'fell out of the carriage' (predicated on an IEM basis) etc. are excluded. But this categorial restriction need involve no infringement of any weak Generality Constraint. Only when we strengthen the Constraint, and introduce the notion of fundamental identification, is there a violation - unsurprisingly, as these are Greek ideas. This, at least, is how it appears. But Trojan deficiencies may be revealed when we consider whether it is required for the subject to have a fundamental identification of himself if he is to grasp the undecidable propositions.

Let us summarise the discussion so far. Evans seems to be saying: 'Trojanism says that I cannot grasp the undecidable propositions, and thereby flouts the Generality Constraint'. The reply so far is: 'Trojanism doesn't want to deny that I can grasp these propositions, and wants to accept a weak Generality Constraint. Its account of the undecidable propositions does not flout the Generality Constraint'. The final instalment is: 'Trojanism may not be able to give a satisfactory account of the undecidable propositions, and fundamental flaws in the Trojan position may thereby be revealed. But this deficiency has nothing to do with the requirements of the Generality Constraint per se'.

Now, our final instalment is the following: It is not adequate to treat the undecidable propositions as merely a species of non-IEM proposition (though they are that) for the following reason. In order to grasp them, the Greek will contend, one needs to have a fundamental Idea of the self (laconically, one needs to view oneself as a person among persons). In the case of 1st-person decidable non-IEM propositions, it is not immediately clear that my method of verification is one others would use in verifying the appropriate 3rd-person transformation. But in the case of the undecidable propositions, there is only the 3rd-personal method of verification - so for me to grasp what it would be for myself to satisfy the predicate 'was breast-fed' it is necessary for me to grasp what it

is for someone to satisfy the predicate. As Evans writes, 'one's general understanding of what it is for a person to satisfy the predicates [concerned] ... [must] be coupled with one's Idea of oneself to yield an understanding of [the undecidable propositions]'

The Greek may continue his argument as follows: Not only do the undecidable propositions differ from the other non-IEM propositions in requiring a fundamental Idea of the self - further, they are undoubtedly required for self-consciousness. For they do, in many cases, involve psychological, and not bodily self-predication, and so the favoured Trojan move for dealing with non-IEM propositions (that they refer to a body), cannot succeed. Thus, in order to grasp a small but significant range of psychological judgements in the 1st-person (those which refer to a state in which consciousness is absent), one must be able to view oneself as a person, an element of the objective order.¹

The first part of this final response looks very plausible - which means that Evans was on to something with his original point about the undecidable propositions, although his presentation was unclear. The second part cannot really be assessed without going more fully into the conditions for self-consciousness, and the Trojan account thereof - which due to the quantity of material on other aspects of the Homeric Opposition, cannot now be discussed at the length it requires.

¹ There are also decidable non-IEM self-ascriptions which are psychological - see Evans' example of where it makes sense to ask 'Someone is feeling a piece of cloth, but is it I?' (ibid. pp. 219-20).

2.4) Anscombe's Tank Argument (I): The Self-Reference of 'I'

(i) Introductory

Anscombe's [1975] has as its most beguiling justification for Trojanism that which I term the 'Tank Argument' (so-called because it involves imagining oneself to be in a 'sensory-deprivation' tank¹). However, various considerations intended to undermine the idea that 'I' refers are introduced by Anscombe. She begins by discussing the view that the role of 'I' is to effect self-reference. The notion of 'self-reference' is primarily a topic for Chapter 3, and the ideas presented here will be taken up at that point. However, a preliminary discussion seems in order since the dialectic of Anscombe's article might then be more clearly understood (if 'dialectic' is not too grandiose a term to use in this connection).

Anscombe does in fact introduce a structure for her arguments - though many of the latter are of interest outside of it. It can be divided into two parts:

- (i) We seem to need a sense to be specified for the quasi-name 'I' (ibid. p. 48); irrespective of which category of singular term it belongs to. But no account of its sense seems adequate.
- (ii) If we stick at the level of reference, it seems the reference of 'I' is especially guaranteed. But the idea of such a variety of reference leads to incoherence (the Tank Argument), so by reductio ad absurdum we may conclude that 'I' does not refer.

The first part may look as unpromising as its crudely Fregean

¹ And not to be confused with Winston Churchill's Tank Argument of 1915.

assumption: that there must be some descriptive formulation of the condition of application of any singular term 'm' that ensures its unique specification. (Confuted by Kripke, Evans [1973] et al.) It is in this part that the account of self-reference fits; but the account is indeed of independent interest.

By 'self-reference', I mean the principle that in self-conscious thought or utterance, the subject (the person who has the thought / makes the utterance) is identical with the object (the person the thought / utterance is about). It should not be confused - as it is by Nozick in his [1981] (see pp. 113-14) - with what I elsewhere term 'auto-reference', featuring in e.g. 'This sentence is in English'. Anscombe shows that to present 'I' merely as a device of self-reference is not sufficient to characterise its sense (or meaning), for the following reason: '"I" is the word each one uses in speaking of her/himself' will not distinguish the 1st-person pronoun from devices that can effect unwitting self-reference (*ibid.* p. 47); nor from 'A' (a self-referential term not expressive of self-consciousness - *ibid.* pp. 49-50 and see Ch. 3 pp. 136-39). An example of the former, elaborated from a work of Jane Austen, is the following: Suppose that before her sudden acquisition of self-knowledge, Emma Woodhouse remarks 'The young lady who has won the tender affection of dear Mr. Knightley is fortunate indeed'. Emma in fact succeeds in referring to herself, but not in the appropriate self-conscious way, since she intends to refer and believes she has referred to her protege, Harriet Smith. Another example is that of Oedipus (see p. 113). (It is arguably a central aspect of dramatic irony that a character self-refers unwittingly because of imperfect self-knowledge.)

(ii) Qualifying the self-reference principle

If the initial account is qualified by 'knowingly and intentionally' (*ibid.* p. 47), is it possible to avoid insufficiency without falling into circularity? If we say '"I" is the word each one uses when s/he knowingly and intentionally speaks of her/himself', will the qualification serve to provide an account that distinguishes 'I' from 'A', but which does not include a form of the reflexive

pronoun 'her/himself' requiring explanation in terms of 'I'?¹
 (We are interested in any 'A'-like term, of course, like one's own name, or some definite description that is true of oneself. We want an account that rules out those cases of self-reference where the speaker doesn't realise he is referring to himself by means of such an expression. There need be no worry about cases where he does, e.g. in one speech by Othello; for there, 'Othello', when used instead of 'I', is expressive of self-consciousness.)

Anscombe denies that the qualification 'knowingly and intentionally' has this desired effect. She poses the dilemma implicit in the foregoing: either 'himself' or 'herself' is the ordinary ('direct') reflexive, in which case the account is insufficient, or it is a special ('indirect') reflexive, in which case the account is circular. Take the former interpretation first: Doesn't Emma knowingly and intentionally refer to herself when she says 'The young lady who has won the tender affection of Mr. Knightley is fortunate indeed'? Anscombe considers the objection that since ascription of intention creates referentially opaque contexts, we might be right to say that X intends to refer to an object, specified under some description, which is in fact identical with her/himself, whilst not intending to refer to her/himself (due to ignorance of the identity). But this is incorrect, she contends, if the reflexive pronoun is the ordinary one, since the use of that pronoun gives no indication of the 'conception' of the object possessed by the speaker. Emma does indeed, then, knowingly and intentionally refer to herself in the utterance cited, if the reflexive is the ordinary one. For the use of the ordinary reflexive 'herself' carries no implication that the speaker self-consciously self-refers - i.e. self-refers in that way incompatible with ignorance that one is self-referring. Thus, if 'her/himself' is the ordinary, direct reflexive, the 'knowingly and intentionally' proviso will not be adequate to rule out unwitting self-reference. (The behaviour of the direct reflexive is well-illustrated in Altham's [1979] (p. 28), where he displays the following equivalence:

¹ This amounts to saying that the reflexive pronoun has only one sense, and that 'knowingly and intentionally' serves to qualify the self-reference as self-conscious.

Reginald hurt himself [direct reflexive] falling downstairs
 = Reginald hurt Reginald falling downstairs,
 and so on for any term co-referential with 'Reginald'.)

Anscombe goes on to imply that if we want a reflexive pronoun which will specify the relevant conception of the object, that is, a self-conscious conception, we will have to treat the reflexive as a 'special' one, viz. the 'indirect' reflexive. And this is what we would naturally do, I think; it must be conversationally implied by '... spoke of / referred to her/himself', a fortiori by '... knowingly and intentionally referred to her/himself', that the reported speaker had the appropriate 1st-personal conception. But in this case, as Anscombe says, circularity is introduced; for it is definitive of the indirect reflexive, on Anscombe's interpretation, that it express the speaker's 1st-personal conception of her/himself. Yet it is this that we are trying to explain.

(iii) 'A'- and 'I'-reflexives; circularity

The distinction between the two senses of 'her/himself' is not quite right, however, although Anscombe's essential point is sound. It is worth explaining why. It is because the '1st-personal' conversational implication of the indirect reflexive can be cancelled or discarded, as in e.g. 'Emma, without realising it, said that she herself was fortunate indeed'.

The point is, that the reflexive of indirect speech is not on all occasions the 'special' reflexive that Anscombe wants. Let us call the latter, which implies the speaker's self-conscious conception, the 'I'-reflexive, and the 'ordinary' self-reflexive, which carries no such implication, the 'A'-reflexive (after Anscombe's distinction between 'I' and 'A'; 'A' also standing for 'Audience' - see below). In the Emma example, 'herself' is the indirect but 'A'-reflexive. The 'A'-reflexive does only what Anscombe says the direct reflexive does. It specifies for the audience the object which the speaker referred to, but offers no conception under which the speaker thought of it, and under which some co-referential substitutions may fail. It is the defining characteristic of the 'A'-reflexive that it tells us that the

speaker referred to the object that s/he in fact is, without telling us whether or not s/he knew that s/he was identical with that object.

Now our original question was, will the 'knowingly and intentionally' qualification avoid both insufficiency and circularity? Anscombe's answer should be rephrased: Either the 'her/himself' will be the 'A'-reflexive (though indirect) in which case the account is insufficient. Or it will be the 'I'-reflexive (always indirect), in which case it is circular. Even the 'knowingly and intentionally' qualification will not guarantee that the reflexive is the 'I'-reflexive. This is because an interpretation denying the conversational implication 'himself' = 'I'-reflexive is always possible, such that the following analysis is correct:

Smith knowingly and intentionally referred to himself
= Smith referred to someone and that someone is himself

This analysis is compatible with Smith's ignorance of the fact that he referred to himself ('I'-reflexive). The following analysis ensures that the reflexive is the 'I'-reflexive but at the price of circularity:

Smith knowingly and intentionally referred to himself
= Smith referred to someone, knowing and intending that person to be himself.

The circularity comes out when we ask 'What does Smith know and intend?' For the oratio recta formulation of his knowledge or intention must use the 1st-person. What he knows is: 'I am the person referred to'.

The circularity thus consists in the fact that the subject's knowledge can be expressed only in terms of the 1st-person, when the reflexive is interpreted in such a way as to rule out ignorant or unintentional self-reference. The account of 'I' (that is, the 1st-person) suggested by the foregoing is as follows:

'I' is the word each one uses (a) to refer to some person,

(b) knowing and intending that person to be himself.

Anscombe views accounts of 'I' which attempt to present it as a device of self-reference, as either insufficient or circular, therefore. Is she then entitled to dismiss the self-referential aspect? It has been nicely put by Evans, that the requirement that a self-conscious thought be self-consciously about the subject is not by itself an adequate basis for an analysis of self-consciousness (because of circularity), but it is nonetheless a principle which must be respected ([1982] p. 259). Anscombe would reply that the 'self-reference principle' is indeed not an adequate basis for self-consciousness (as shown in the discussion of the indirect reflexive); and this for the very good reason, on the Trojan view, that it is false. There are thoughts which on this view are governed by some version of the principle: viz. those of the 'Emma' variety, and 'A'-thoughts. In such cases, the living human being which has the thought is identical to the object towards which the thought is directed. But this is to down-grade the principle, of course; for thoughts of the sort I attributed to Emma, are not, after all, very interesting, and 'A'-thoughts occur solely in thought-experiment. Thus for Anscombe there may be and indeed is a relation between the subject-term of the 'I'-thought or utterance and the living human being who is the thinker or speaker; but it is not that of object-direction or reference. Nor is there anything else to which the subject-term can stand in such a relation (see pp. 79-80). '"I" is the word each one uses when s/he knowingly and intentionally speaks of herself (or himself)' is correct only insofar as it does not imply the reference-relation; and is in any case circular. (A 'misleading platitude' therefore.) Assessment of these views awaits the denouement of part 1 of the Homeric Opposition. The independent issue of the self-reference principle and its alleged circularity are treated in Chapter 3.

(iv) Discussion of non-self-conscious self-reference

The next part of Anscombe's account is concerned with elucidation of the nature of self-consciousness - and as such is also discussed more fully in Chapter 3. However, it fits in with the preceding anti-Greek dialectic, in the following way. Anscombe

compares 'I' with a self-referential term 'A' which, unlike 'I', does not express self-consciousness.¹ Another 'misleading platitude' is here undermined: that self-consciousness is consciousness of the object that one is. Again, the Evans-style objection is in order: the fact that consciousness of the object that one is can be possessed by non-self-conscious beings and is thus not an adequate basis for an analysis of self-consciousness, does not mean that it is not a necessary component of self-consciousness. For Anscombe, on the other hand, the inadequacies of such platitudes are a symptom of that which she intends to demonstrate: that 'I' is not a referring expression. The discussion of 'A' fits into this overall strategy in the following way: it shows that one 'trivial' explanation of why 'I' is not colloquially a proper name (that it won't do the work of a personal proper name because everyone has it and uses it only to self-refer) will not do, and that there must be a deeper reason ('I' is not any kind of referring expression; it is not colloquially a proper name, nor is it a 'logically proper name' - a singular term).

¹ 'A'-users lack the capacity to make IEM-judgements; in their 'A'-utterances they 'mean to speak of a certain human being, one who falls under [their] observation in a rather special way'. (See below, pp. 136-39.)

2.5) Anscombe's Tank Argument (II): Presentation of the Argument

(i) Prolegomenon

We have by this point dealt with the self-referential proposal for the sense of 'I', and the amendment of it, which Anscombe makes in part (i) of her destructive account of 'I' (p. 57 above). She has a further proposal, and with it switches to the 'mode of presentation' formulation of Frege's notion of 'sense', but in fact thereafter employs a more limited, less contentious requirement: 'the use of a name for an object is connected with a conception of that object' (ibid. p. 51); a sortal term must be supplied for each putative proper name (ibid. p. 52). This last is associated less with Frege's notion of sense, more with his insistence that a name must have associated with it a criterion of identity (see Dummett [1981] pp. 73-6).

The new proposal is introduced thus:

If things are, rather than having, selves, then a self is something, for example a human being, in a special aspect which he has as soon as he became a 'person'. 'I' will then be the name used by each one only for himself (this is a direct reflexive) and precisely in that aspect (ibid. p. 51).

When the aspect is explained, we may have that "way of being given" of an object' which is associated with its name, Anscombe notes. But this proposal is nonsensical, she maintains. Her objections seem to be:

1. One might refer to oneself in the wrong way, on this account - the old point about the inadequacy of the 'direct' (or as I must now put it, the 'A'-) reflexive.
2. One might not refer to oneself at all: it would be a question what guaranteed one got hold of the right self (ibid. p. 51).

Now this reaction to the proposal may seem unduly vituperative. It seems harmless and salutary to link the meaning of 'self' with that of 'person', and to define them in terms of a special aspect a human being may have. But Anscombe is still after a non-circular account of 'I'; and the talk of aspects will not, when added to the bare self-referential account employing the 'A'-reflexive, make that account adequate. It will not stand proxy for the circularity-inducing requirement that the speaker self-consciously self-refer (or however we are to express this vertiginous thought). Nevertheless, once again the riposte to Anscombe is: insufficiency must not be equated with falsity. Will her second objection sustain the latter charge? As it stands, it is an objection to the unpromising idea that one has, rather than is, a self. But it can be developed into a criticism of the aspect account too; for that account leads to a transgression of the Doctrine of the Secure Conception of the 'I' (as I will call it). Such a conception is 'secure' only if it ensures that the reference of 'I' is guaranteed (in ways to be specified). It will be recalled that Anscombe uses the phrase 'conception of the object' in framing the sortal requirement (p. 64). The present doctrine thus amounts to the idea that whatever sortal term is placed in apposition to the quasi-name 'I', it must be such as to ensure 'guaranteed reference' for 'I'. The second objection to the aspect account then becomes: if I am trying to pick out a self by means of some expression, it is a possibility that I pick out the wrong one. But with 'I', the object I pick out by means of it must be identical with the object I intend to pick out. (And a fortiori, I cannot fail to pick out anything).

This is to anticipate somewhat. For Anscombe's immediate and main concern is with the new proposal seen as an expression of 'Cartesian Ego theory'; the self seen as an imponderable something one has, rather than something one is. This unpromising construal is not worth discussing.

The final component of the first part of Anscombe's negative account of 'I' is the introduction of the notion of 'guaranteed reference', explored more fully in the second part. The Doctrine of the Secure Conception of the 'I' is implied in the following passage:

... what conception can be suggested, other than that of thinking, the thinking of the 'I'-thought, which secures this guarantee against reference-failure? It may be very well to describe what selves are; but if I do not know that I am a self, then I cannot mean a self by 'I' (ibid. p. 55).

Now, what can it mean to talk of a 'conception' securing guaranteed reference? (I am not disputing that 'I' does have guaranteed reference in the sense stated on pp. 12-13). First we need an appropriate sortal from the 'conception' 'thinking', viz. undramatically, 'thinking thing' ('I am a thinking thing' is equivalent to 'I am thinking'). Now, why will this sortal alone guarantee reference, where 'self' fails?

(ii) The Tank Argument of the first part of Anscombe's dialectic

Anscombe's answer is cryptic, making use of that component of the guaranteed reference of 'I' by which it is required that the referent of 'I' be really present to consciousness (ibid. p. 55). But we can construct a defence of the view that 'thinking thing' is the only conception that will serve, from the Tank Argument, which is ostensibly an argument in part (ii) of the destructive account, at the level of reference. So far, particular accounts of the sense of 'I' have been discussed and found wanting by Anscombe; now what is offered is a quite general argument to the conclusion that no coherent account of the sense of 'I' can be provided. I merely present the 'Tank Argument' now. A slightly extended version is offered in the course of discussion of part (ii) of Anscombe's destructive account and the latter version contains all the difficulties of the following (plus a few more!) - these are idealisations of her compressed account:

- (1) If 'I' is a referring expression, then each serious and comprehending use of 'I' is guaranteed to have successful reference, where success consists in (a) referring to an object that (b) one intends to refer to (i.e. the object referred to fits one's conception of it).

(2) 'I' refers.

so (3) 'I' has guaranteed reference.

For the subject in a sensory deprivation tank (see Anscombe ibid. p. 58):

(4) It is certain that the conditions for the successful reference of 'I' are met in the Tank.

(5) It is not certain that the referent is embodied.

so (6) That the referent is embodied is not one of the conditions for successful reference.

so (7) It is no part of any (secure) conception of the referent that it be embodied (i.e. 'thinking thing' is the only sortal term that secures guaranteed reference).

but (8) Such a conception is internally incoherent.

so (9) We can attribute no sense to 'I' if we are to respect its guaranteed reference.

Now just what is wrong with the conception 'thinking thing'? The problem is, that the guarantee that one refers to the object one intended to, is purchased at the cost of the impossibility of intending to refer to the same object that one has referred to, on a previous occasion. Anscombe comments that we have been driven to Descartes' notion of a res cogitans and writes:

... if 'I' is a referring expression, then Descartes was right about what the referent was. His position has, however, the intolerable difficulty of requiring an identification of the same referent in different 'I'-thoughts ... (ibid. p. 58).

The reason why 'thinking thing' secures guaranteed reference is that, in contrast to the other sortal terms considered, it has no criteria of identity that could be misapplied; so there could be no question of intending to apply 'I', a name for a token of the type,

to the same object one has applied it to on a previous occasion, and failing. Thinking things cannot be the object of more than one act of reference each; their 'identity' persists only for the duration of the 'thought-episode', which is no time at all. For what possible criteria of identity could there be that would enable a 'thinking thing' to be re-identified, if its only essential feature is the particular thought-content?

In fact, it is worse than this. It is not clear that a 'thinking thing' can be the unitary object even of one act of reference. Not only must 'I' refer to something the existence of which is instantaneous, but it might refer to several things on a particular occasion for all one knows. For 'thinking thing' has no criteria of individuation either. Anscombe writes:

How do I know that 'I' is not ten thinkers thinking in unison? Or perhaps not quite succeeding. That might account for the confusion of thought which I sometimes feel (ibid. p. 58).

According to Anscombe, then, 'thinking thing' is a pseudo-conception for a non-object.

(iii) A fallacious false start

I will now present the Tank Argument as it fits in to part (ii) of the destructive account. As presented, the argument relies on aspect (1) of guaranteed reference - the guaranteed existence of the object meant by the 'I'-user.¹ This argument is an extended version of that given on pp. 66-67.

The Tank Argument, then, runs as follows:

- (1) If 'I' is a referring expression, then each serious and comprehending use of 'I' is guaranteed to have successful reference, where success consists in (a) referring to an object that (b) one intends to refer to (i.e. the object

¹ See section on guaranteed reference, pp. 12-14.

referred to fits one's conception of it).

(2) 'I' is a referring expression.

so (3) Each serious and comprehending use of 'I' is guaranteed to have successful reference.

Now for the subject in a sensory deprivation tank:

(4) It is certain that the conditions for the successful reference of 'I' are met in the Tank (assuming it is plausible for the subject to have an epistemic attitude towards this technical truth).

(5) It is not certain that the referent is embodied.

so (6) That the referent is embodied is not one of the conditions for successful reference.

so (7) It is no part of any (secure) conception of the referent that it be embodied ('thinking thing' is the only sortal term that secures guaranteed reference).

but (8) Such a conception is internally incoherent.

so (9) ($\neg(3)$)

and(10) ($\neg(2)$)

It is regrettable that this argument, on close scrutiny, requires an additional premise to avoid fallaciousness. Premise (4) requires strengthening in the following way: not only must it be certain for the subject that the conditions for successful reference are met, but he must further know what the conditions are. He must know all truths of the form:

'It is a condition for C that q'.

Thus, 'that the referent is embodied' may be one of the conditions

for successful reference of which I know that, whatever they are, they are satisfied; but if I do not know it is one of the conditions, I may not be certain of its truth. Analogously, a person even more ignorant of the laws of Newtonian mechanics than the present writer, observes that his car is stationary. So he is certain that the conditions for its being stationary are met. But due to the defects of a classical education, he is not certain that the forces acting upon the car are equal and opposite (he just doesn't know about such arcane matters).

It seems that the mistake in this version of the Tank Argument, then, is the following: 'There is something of which I have no guarantee (that I am embodied). This cannot figure as a condition of something of which I do have a guarantee (that every act of reference involving 'I' is successful). (So 'I' doesn't refer to something essentially embodied.)' But we have seen that this first step is invalid; for in addition to knowing that the conditions obtain, I need to know, of each of the conditions, that it is a condition. To repeat, knowing that the conditions for successful reference (whatever they are) obtain, is compatible with ignorance that one of the conditions, specified in 'logically isolated' terms as 'that the subject is embodied', obtains; I can be ignorant of this truth, if such it is, only if I do not know that it is one of the above-mentioned conditions (if it is such). So this version of the Tank Argument will not of itself achieve for Anscombe the conclusion that I am not essentially embodied ('I' does not refer to something essentially embodied).

(iv) Climax of the Anscombian dialectic

The defective argument can be amended, however, to avoid the fallacy and to present three alternatives (4, 4' and 4'' below) for negation in contraposition. This is the definitive Tank Argument; for those who, unlike the present writer, are not afflicted with logical dyslexia, the labelling of assumptions and inferences may help understanding. It should be emphasised that, as in previous versions, the argument is to be administered to himself by the subject in the Tank:

- (1) It is certain that: if 'I' is a referring expression, then each serious and comprehending use of 'I' is guaranteed to have successful reference. [Ass.]
- (2) It is certain that 'I' is a referring expression. [Ass.]
- So (1,2) (3) It is certain that each serious and comprehending use of 'I' is guaranteed to have successful reference. [from 1,2]
- So (1,2) (4) It is certain that the conditions for the successful reference of any genuine 'I'-thought are met in the Tank. [from 3]
- (4') It is certain that embodiment is necessary for a successful referential use of 'I'. [Ass.]
- (4'') It is certain that 'I'm missing the 2.30 at Goodwood', thought by myself in the Tank, is a genuine 'I'-thought. [Ass.]
- So (1,2, 4',4'') (5) It is certain that when I think 'I'm missing the 2.30 at Goodwood' in the Tank, I am embodied. [from 4, 4',4'']
- But (6) (~~=5~~) It is not certain that when I think 'I'm missing the 2.30 ...' in the Tank, I am embodied. [Ass.]
- So (4', 4'',6) (7) (~~=4~~) It is not certain that the conditions for successful reference for any genuine 'I'-thought of the sort which could figure in premise (6) are met in the Tank. [from 6, 4',4'']

This rearrangement serves to focus attention on the main point at issue, whether the subject's doubt about embodiment is possible. It is this point which affords Trojanism its strongest intuitive

purchase. (It is true that, on the second option for the Greeks that I will outline, it is accepted that the doubt is possible - but no clear account of the thought involved (said not to be a genuine 'I'-thought) is given.) The rearrangement also avoids the aforementioned fallacy. It should be noted that two things are presupposed by the argument if it is to yield a Trojan conclusion.

First, that certainty is transitive across logical consequence; viz. that the following principle holds:

$$\frac{C(\Gamma), C(\Gamma \vdash q)}{Cq}$$

Second, that the idea of doubting the existence of one's body, in the absence of current information about it, is intelligible and unproblematic. The latter is not uncontroversial. Finally, is the conclusion as it stands in fact Trojan? After all, what is claimed is just that it is not certain that the conditions for successful reference, for any of the 'I'-thoughts in question, are met in the Tank. The argument requires supplementation, then, to take it beyond the original Anscombe conclusion, as follows:

(8) It is a truth knowable a priori, if it [Ass.]
is a truth at all, that 'I' is a referring
expression.

(9) For any a priori truth, it must be [Ass.]
possible for one to be certain of it.

(4', 4'', 6) (10) In the Tank, it is not certain that the [=7]
conditions for successful reference of
any genuine 'I'-thought which could
figure in (6) are met.

So (4', (11) It is not certain that each serious and [from 10]
4'', 6) comprehending use of 'I' (in the sort of
thought which could figure in (6)) is
guaranteed to have successful reference.

So (1,4', (12) It is not certain that 'I' is a referr- [from 1,11]
4'',6) ing expression (in the sort of thought
which could figure in (6)).

So (1,4', (13) It is not the case that 'I' is a [from 8,9,12]
4'',6, referring expression (in the sort of
8,9) thought which could figure in (6)).

2.6) Anscombe's Tank Argument (III): Assessment

(i) Three Greek options

So the conclusion is that in a range of uses at least, 'I' is not a referring expression. This range corresponds to the range of 'I'-thoughts which the subject can entertain consistently with doubting the existence of his body. The non-referring uses will thus be those in 'psychological' self-ascriptions; which, if one had not thought too deeply about the matter, one might think were the IEM-uses. (Of course, as we have already shown, there are physical self-ascriptions which are IEM.)

What is happening at this point in our discussion is of the first importance. What is the thought expressed in (7) (p. 71)? It is that maybe my 'I'-thought has no object (to dismiss the alternative variety of unsuccessful reference: that I end up referring to an object other than the one I intended to). Now the Greek can react to the argument in two ways. On the first option, both sides will agree that the thought that my 'I'-thought has no object is implied by the thought that maybe I am not embodied (on the assumption that it is certain both that embodiment is necessary for successful reference and that the contemporaneous 'I'-thought in (4'') is a coherent one). It is in a sense equivalent to it. Anscombe writes (in a vein that, as will become apparent, I am not sure is a Trojan one):

[In the Tank] I should perhaps believe that there is ... a body. But the possibility will perhaps strike me that there is none. That is, the possibility that there is then nothing that I am (ibid. p. 62).

But the thought 'Maybe my 'I'-thought has no object / I am nothing' seems impossibly difficult to entertain. Like the brain in a vat, the subject who believes himself to be disembodied is trying to work with, as Evans puts it, an inadequate Idea of himself. (But of course he is simply deluded, whereas with the unfortunate Brain, not

even psychotherapy could help.) The fact that the doubt about embodiment amounts to the doubt about object-direction, and that the latter can be represented as the absurd doubt about one's own existence, leads the Greek to contend that the superficial plausibility of the former hides a deep unintelligibility. Still, the idea that the doubt about embodiment can be entertained is an appealing one, and if he denies it, we need from the Greek a diagnosis of the error. A possible diagnosis, suggested by Evans' discussion, is as follows: 'It is just the old mistake of thinking that what happens sometimes can happen always. In the Tank, the subject is temporarily deprived of one kind of information about himself which normally forms the basis of his self-conscious thought. But we allow that even in this situation, he has 'I'-thoughts. So why should it not be the case that all self-conscious thought involves no more than these 'I'-thoughts do? This gives us the idea that there could be a subject who never received information about his body or anything else, who in fact has no body to receive information about. And this is the content of the doubt in the Tank. But it is a requirement on self-conscious thought that the subject has the capacity to locate himself as an element in the objective order¹; and though we can credit this capacity to a subject for whom circumstances prevent its exercise, though it once was exercised, we cannot credit it to a subject which ex hypothesi could never exercise it. Disembodiment perhaps makes sense if, à la Strawson [1959], we make it essential that the subject has been embodied; but such a subject will not be the essentially thinking subject of Anscombe's conception'.

This first option, then, consists in a refusal to countenance the doubt about embodiment, on the grounds that it involves a doubt that one obviously cannot entertain, a doubt that amounts to a doubt about one's own existence. The subject of an 'I'-thought cannot think of himself as nothing and nowhere. So by reductio ad absurdum, it must be held that it is certain that I am embodied in the Tank.

¹ Alternatively one could say: self-consciousness essentially involves ways of knowing things about the object that one is.

The second Greek option, in contrast, consists in an acceptance of the doubt about embodiment, and an acceptance of one of the alternative consequences. The Greek accepts the plausibility of (6) and (4''). The effect of the stratagem is not much different from denying (6) straight off. But it means that (4), (4') and (4'') are seen clearly as a trilemma, with the two chief protagonists, and a third disputant, each selecting a different premise for negation. The Trojan response discussed shortly, is as presented in the argument; the denial of premise (4), that it is certain that 'I' refers in any coherent 'I'-thought in the Tank. The Cartesian, the third disputant and too weak as a Mediterranean power to be afforded a full part in the Homeric scenario, will choose to deny (4'), that it is certain that embodiment is necessary for a successful referential use of 'I'. Thus one can say that the Cartesian view contains elements of the views of the Homeric opponents; the Trojan attitude towards embodiment, and the Greek attitude towards reference. The result is unsatisfactory, as has been noted in the discussion of the concept of a thinking thing; viz. reference to something imponderable. It should be noted, incidentally, that premise (4') is in fact acceptable to Trojanism - because otiose from that viewpoint. It may be the case that embodiment is necessary for successful reference by means of 'I'; but since 'I' does not in the cases in question refer, successful reference is no part of its use. (In fact, of course, there is a referring use of 'I' - the 'I'-as-object use featured in bodily self-ascription. That successful reference in this case obviously requires embodiment is a further justification for (4') on the Trojan account.)

The Greek response on this second option is thus to claim that if it is admitted that it is doubtful that I am embodied, then it will follow that it is doubtful that my putative 'I'-thoughts are genuine 'I'-thoughts - for the reasons advanced before, equating the doubt about embodiment with a doubt about object-direction. On this option, however, the doubt is allowed some intelligibility. But as it is not allowed to be an 'I'-thought, it is hard to say what structure it could possibly have. For this reason, the second option does not seem very plausible.

On the Trojan account, in contrast, though the doubt about embodiment amounts to the doubt about object-direction, the latter is not interpreted as the doubt about one's existence. (Though it is not easy here to make sense of Anscombe's contention that I therefore doubt whether there is then anything that I am. Her view seems to be that 'I'-thoughts necessarily stand in some relation other than that of reference, to a living human being.) Rather, one's doubt about object-direction amounts to an insight into the objectless nature of self-conscious thought.

What is the current state of play between the Homeric opponents then? The only viable Greek response so far, is to question the possibility of the doubt about embodiment. This is not a very tractable issue. But in fact there is a third Greek option also.

The third Greek option is like the second option but negates premiss (4) about guaranteed reference instead of (4'') about the genuineness of the 'I'-thought. But then the Tank Argument, apparently Trojan, could be represented as compatible with Hellenism. Curious - but there is an analogy to this option in the account by Evans in his [1982] of 'the perennial nightmare' (perennial in the imaginings of philosophers), the brain in a vat horror-story. The scenario is that a human brain is taken at birth and placed in a vat, where it is stimulated by scientists in such a way that it hallucinates normal human experiences and is then allowed a 'normal' cognitive development. The subject makes the usual range of self-ascriptions - all founded on the delusion that it is a normal embodied person. The question arises: how ought the subject to think of itself? Or, how, apprised of its true situation (if that is the correct term!), can it think of itself? As the subject in the sensory-deprivation tank believes he can think of himself, so the brain must attempt to think of itself - if, with Evans, we deny that the latter can identify itself with a brain in a vat (ibid. p. 251).

What, then, will be the structure of these thoughts that, as Evans says, 'fail to net any object at all?' On the Trojan account, it may be said that there is no problem: the brain's thoughts share the structure of 'I'-thoughts, and it is not the case that there is a body which the enquirer might be misled into thinking had a role

to play as referent. Anscombe would not be quite happy with this response, for the reason mentioned above - that there is a relation between the 'I'-thought and the living human being who has it, albeit not that of reference. On the Greek account, there are two options, corresponding to the second and third options above. The former will then be that the brain does not have the 'I'-thoughts, such thoughts being guaranteed of reference. Its thoughts have the surface structure of 'I'-thoughts (if one can attribute such a structure to thought), but in reality are 'Je'-thoughts - where 'Je' is a non-referring term signifying that there is a subject that is not an object. Obviously this is just a label! It will not answer the query about the nature of the doubt about embodiment when, on the second Greek option, that doubt is allowed to be intelligible. To allude to a recent debate, it might be held that the brain does not have 'I'-thoughts because such thoughts are about an actually existing object, and could not exist without that very object's existing - that is, they are de re thoughts (see Woodfield ed. [1982] p. v).

The option favoured by Evans is the latter: these thoughts are 'I'-thoughts featuring a non-referring use of 'I' (so 'I'-thoughts are not de re). Thus he claims that 'the Cartesian assumption that [our ordinary thoughts about ourselves] are always guaranteed to have an object cannot be sustained' (op. cit. p. 249). But is this not a Greek assumption also? It need not be. The Homeric opponents hold respectively that 'I' is, in its 'psychological' (or in its IEM uses) a referring (or non-referring) expression. But Strawson himself, the Agamemnon of the dispute, seems to suggest that there is a non-referring use of 'I' - albeit the transcendental use ([1966] p. 166). So it is only at first sight curious that the Tank Argument as presented on pp. 71-72 might have a conclusion acceptable to the Greeks. (The conclusion being that 'I' does not refer in 'psychological' thoughts entertained in the absence of sensory information from one's body.) If it is not certain that I am embodied, then it is not certain that my 'I'-thoughts refer - in the unusual circumstances of Cartesian doubt in the tank. On an option acceptable to Greeks with a taste for outré thought-experiments, the latter thought, instead of leading to a reductio ad absurdum of the former, might be regarded as entertainable.

Although Varieties of Reference suggests this line of thought, Evans himself would not be happy with its application to the Tank Argument, however. He emphasises that a subject's Idea of himself does not require him to have a 'current conception' of himself:

... what is required, in the exceptional circumstances in which the various avenues of self-knowledge are blocked, is that the subject be disposed to accept any information accessible in those ways as germane to the thoughts we regard as manifesting self-consciousness (ibid. p. 249).

Cases where the 'Cartesian assumption' about reference-failure is not sustainable are 'science-fiction' examples - of which the Tank is not one (ibid. pp. 249-55). I would be inclined to agree with this assessment - which amounts to a refusal to allow that the possibility of Cartesian doubt in a sensory-deprivation tank can have any very important consequences.

I have done my best to present the Tank Argument in its strongest possible terms, shorn of any obvious fallacy. Despite the second and third Greek options just canvassed, the first I think locates the key point at issue - the doubt about embodiment. We can take discussion of this issue further by considering the merits of arguments put forward by Kenny in his [1979].

(ii) Kenny on 'The First Person'

In the course of Kenny's discussion, there is some possible confusion over the separateness of the Trojan and Cartesian viewpoints which should be dealt with first. He discusses the Tank Argument, states what he takes to be Anscombe's conclusion, and criticises it:

'I am the thinker of these thoughts' is not a genuine proposition, but 'I am this body' is ... [she claims]. This conclusion I found surprisingly unWittgensteinian: the thinker of these thoughts who is possibly not this person with this body seems uncomfortably similar to a Cartesian ego; and if we allow the conceivability of the notion that we are thus spiritual, it

seems no great matter that when we use 'I' we are not actually referring to any such spirit or self (ibid. p. 13).

It is this line of thought and, connectedly, his concern that in her positive account of self-consciousness, Anscombe is making play with the bête noire of private ostensive definition, that makes Kenny query, only half-ironically I think, whether the article 'marks the conversion of Professor Anscombe from Wittgenstein to Descartes' (ibid p. 13).

Now it is often thought that, since the Trojan denies that 'I' refers to something embodied, he is thereby countenancing the existence of an incorporeal self. It has to be insisted that the Trojan is not committed to any such view (certainly not consciously, nor I think logically). However, it is not clear that this mistaken view is Kenny's; for he endorses the 'no-reference' view and seems concerned rather that it is Anscombe's use of the Tank Argument and her positive account of self-consciousness that may be driving her to the heretical, Cartesian conclusion. This is a different matter which will now be discussed at some length. (Incidentally, this latter account of self-consciousness is an extreme example of Anscombian cryptic opacity, in this respect anticipating, albeit briefly, the sustained and bravura opacity of her paper 'Propositions'¹ (see her [1975] pp. 61-64).)

In his discussion of the Argument, Kenny considers the problematic status of the Tank-victim's doubt about embodiment. Can I have the thought 'Maybe I don't have a body'? Kenny finds the thought incoherent. He has two arguments, it seems. The first is as follows:

- (1) The assertion of an 'I'-thought is verified or falsified with essential respect to the behaviour of the body; this fact is part of the sense of 'I'.

¹ Delivered to packed houses at St. Andrews and elsewhere in 1983.

- (2) The doubt 'Maybe I don't have a body' must violate this rule (in a way to be specified).

So either

- (3) It has no clear public sense.

or (4) It must have a logically private sense belonging to it when it is entertained in thought (which is an incoherent and pernicious supposition).

Kenny believes that Anscombe will have to endorse the incoherent and pernicious (4); but does the sense in which the doubt violates rule (1) compel Anscombe to this position? Does rule (1) indeed make the doubt incoherent? We need to spell out the rule-violation in more detail.

(iii) The argument from pragmatic self-defeat

Let us postulate (if this has not already been done by someone else) a class of pragmatically self-defeating doubts. Contenders are 'Maybe I don't have a body', 'Maybe I'm dreaming' and (more picturesquely) 'Maybe I'm mad'. These are doubts the verbal expression of which, given a certain condition, defeats the doubt. The precise condition varies somewhat from case to case - usually it is that the verbal expression must be a comprehending one.

Now the point to be urged against Kenny is that just because a doubt is pragmatically self-defeating does not mean it is essentially self-defeating, i.e. incoherent or self-contradictory. Let us examine the looser cases of pragmatic self-defeat, then the stricter cases of which the Anscombian doubt is the limiting one. (The former are indeed sufficiently removed from the last-named that discussion of them is chiefly justified by their intrinsic fascination.)

'Maybe I'm mad' is the grossest example from a number of doubts which refer to afflictions the obtaining of which to a decisive extent rules out the victim's awareness of them. It is the grossest because the terms 'madness' and 'insanity' are ill-defined, and frequently figure as merely pejorative expressions lacking in

content - but the doubt is nonetheless a coherent one which features poignantly in the fears of the neurotic and (arguably) the psychotic. If we take 'mad' and 'insane' in their only clear sense - 'mentally ill' - then to the extent that a psychotic patient in a non-acute phase of his illness has awareness of his condition, the doubt is not in fact pragmatically self-defeating (pace certain accounts¹). (Schizophrenic patients in an acute phase do not display any insight into their condition - but this fact is not sufficient to make the doubt pragmatically self-defeating, because of insight that may be displayed after the phase has passed.) The doubt is a well-attested one amongst neurotics, in contrast, where it may be taken as expressing a more colourful range of fears about losing one's grip, 'cracking-up', etc.; in such cases the doubt is not pragmatically self-defeating, but merely false. (The subjects, assisted by the widespread ignorance about mental illness, have made a serious mistake about their condition; which is not to belittle their fears.) But there are related doubts which are pragmatically self-defeating - where the kind of mental incapacity is of a different sort. One example of this sort of doubt would be 'Maybe my memory has been radically impaired as in e.g. Korsakov's Syndrome' - an exotic doubt, but one which acquires some poignancy when one considers the situation of the hapless victim of the condition referred to, and the extent to which he may be able to obtain some glimpse into his state. It is unlikely that any such glimpse is going to satisfy the requirement of comprehending utterance necessary if the doubt is to avoid being pragmatically self-defeating, however; the subject's sense of self-identity will be sufficiently attenuated that he will not be able to articulate what he has lost.² But no Kenny-style argument is needed to show that the

¹ e.g. Kenny [1973] p. 209.

² Proponents of a more expansive attitude toward the concepts of the congenitally blind may be inclined to differ here. But memory is so integral a component of self-identity that in severe cases of Korsakov's Syndrome it is hard to credit the subject with any insight into his condition; or indeed to understand the quality of his experience at all. In his fascinating [1985], Oliver Sacks discusses the case of his patient, the wretched Jimmie R., for whom one's sympathy is

doubt is not merely pragmatically self-defeating, but essentially so; nothing hinges on the matter of the verbal expression of the doubt.

So we need to look to other examples to see how pragmatic and essential self-defeat come apart. The example of dreaming is discussed by Kenny himself in his [1968], where his views seem at least in tension with his strictures against Anscombe. He writes:

I agree with Malcolm against Descartes that one cannot make judgements during dreams. It does not follow, however, that the judgement 'I am dreaming' is senseless. It can never be made truly, but it can be made falsely. To dream that one is dreaming is not to judge that one is dreaming, but a waking man might be persuaded falsely to judge that he is dreaming ([1968] pp. 30-31).

Kenny here concurs that judgement-making requires that one be

the greater in virtue of his cheerful obliviousness of his affliction. When Sacks suggests he has difficulties with his memory, and illustrates why he might not realise this, he replies: "'So that's my problem ... I kinda thought it was. I do find myself forgetting things, once in a while - things that have just happened. The past is clear, though'" (ibid. p. 25). Yet his terror on being shown himself in the mirror, a middle-aged man 30 years on from the point at which his long-term memory ceases and from the past in which he lives, quickly subsides when the mirror is removed - poignantly illustrating the pathologically-conditioned limits of his self-knowledge. Sacks' analogy of his patient's mental life with the flux and movement of the Humean bundle of perceptions without its unifying principle of personal identity seems, if not philosophically, at least imaginatively, not so wide of the mark. (Incidentally, the book is full of philosophically interesting case-studies, most notably 'the man who mistook his wife for a hat' - now the subject of a tedious opera by Michael Nyman - and the female patient who lost her proprioceptive sense, with horrifying consequences.)

conscious, without concluding 'I am dreaming' is senseless; should he not conclude similarly for 'I have no body', whilst maintaining that judgement-making requires embodiment? Plausibly yes - in which case one must find premise (2) of the argument (p. 80) false. From this it may be concluded that, although principle (1) may furnish a counter-argument to Anscombe's, it will not do so via consideration of the Anscombian doubt as pragmatically self-defeating. The charge of illicitly logical privacy will not be made to stick that way, therefore, and is undermined in general by the following considerations. Like Descartes' Cogito, the Tank Argument is an argument one administers to oneself; but it can generally be so administered. It is true that 'Maybe I don't have a body', like 'Maybe I'm dreaming', can be immune from pragmatic self-defeat only when entertained 'privately', if at all. (I discuss the caveat in the next paragraph.) But the fact that the doubts can be expressed publicly, though subject to self-defeat, shows that Wittgensteinian sensibilities about logical privacy need not be offended.

Three points require immediate statement. First, it is indeed difficult to see how Kenny, and anyone persuaded by his account of the doubt about dreaming, can reject the analogy between that and the doubt about embodiment, since he cannot maintain that the latter is pragmatically self-defeating whilst the former is merely always false. Second, though, I presume that Kenny is also wanting to claim that when contingently privately entertained, the latter doubt is pragmatically self-defeating too. But unlike in the case of the verbal expression, this further claim presupposes an anti-Trojan argument and does not furnish one. That is, whilst clearly one must be embodied in order to give a thought verbal expression, certain proponents of Trojanism might claim that this was not necessary merely for one to have the thought - and to assume this is wrong is merely to beg the question against these proponents. (It is not just Trojans who make this claim - so too does Strawson with arguable consistency, in the famous and putatively Hellenistic cameo of the disembodied person ([1959] pp. 115-6) - reiterated with the authority accorded by the then burgeoning doctrine of Strawsonian infallibility in his [1980] p. 270.) An anti-Trojan conclusion does not follow immediately from principle (1). Third, the Anscombian doubt might even be publicly expressed by the subject in the Tank, though undetected by him. (Solitary persons often unthinkingly give

verbal expression to their solitary thoughts, this being a symptom of loneliness or rejection of the world rather than insanity; under-water microphones could monitor any such pronouncements of the Tank-subject, from which a Becket-like monologue might emerge.)

(iv) A better line of argument

The fact that any such sub-aquatic musings are undetected by the subject is one among the many circumstances that, according to Anscombe, make it seem to him as if he might not have a body any more. But Kenny is on a better tack when, in his second argument, he challenges the grounds rather than the coherence of the doubt:

(1) It is not by sensory experience that I know I have a body.

So (2) Lack of sensory experience in the Tank cannot ground a doubt about whether I have a body.

Claim (1) might be made out by showing that 'I have a body' is a proposition without grounds - the sort of idea Wittgenstein is concerned with in On Certainty.¹ To endorse Kenny's argument here, is to say that the special circumstances of the Tank do not afford a rationale for a particular doubt about embodiment above and beyond that of Cartesian sceptical doubt in general. But in making this an effective objection, Kenny is not exposing any obviously Cartesian inclination in Anscombe's account in the direction of privacy of experience.

He is, however, exposing the key issue in the assessment of the Tank Argument. It is the plausibility of his claim (1) which leads one to feel that Anscombe's doubt about embodiment is, at best, ill-grounded. Further work obviously needs to be done but I

¹ See e.g. para 387:

Someone might ask me: "How certain are you that ... that is your foot?" And the answer ... might be ... "I can't doubt it". And [this] answer would make sense even without any grounds.

close by recording my suspicion that, even on the best presentation of the Tank Argument (which I think I have given), it will succumb to a Kenny-style objection of the above sort. The second and third Greek options of pp. 74-78 are therefore misguided.

2.7) Trojan Arguments from IEM (I): The Unique Guarantee for 'I'

- (i) 'No unnoticed substitution' of the object of self-reference is not the 'unique guarantee', and is empty

It was stated in Chapter 1 (p. 25) that IEM for 'I', as opposed to other indexicals, extends over time, and this seems to indicate a persisting subject statements about which are immune to error through misidentification. Before examining the implications of this immunity from which the Trojan seeks advantage, we must be clear exactly what the immunity is. We need to distinguish two kinds of guarantee about identification over time. The first, perhaps because it is a reflection of 'guaranteed reference' (discussed in Chapter 1), turns out to yield no substantial consequences. It is the second, the genuinely IEM guarantee, which will later form the basis for the important Trojan arguments in section (2.8).

Trojans have attempted to base a case on each of these phenomena, in fact. (If the patient reader has been thinking 'Oh no, not more about IEM', he should be assured that things get rather more exciting from this point onwards.) The weaker, though initially plausible case, based on the first phenomenon, concerns what Anscombe terms 'no unnoticed substitution'. It is met by Evans' argument in *ibid.* p. 214; and this does not involve IEM as we shall see. There is a second case, which does involve IEM; of this, Evans' remarks on pp. 214-5 constitute an inadequate denial, whilst those on p. 218 do not even acknowledge it. However, both cases do involve a guarantee about identification over time in some sense; and we need to show why the triviality of the first case does not affect the substantiality of the second. Matters are not helped by the fact that in Evans' discussion on pp. 213-5, the two cases are run together and jointly diagnosed as expressions of the same error. It is no simple mistake Evans makes here. In the first, the Trojan derives his case simply by presupposing in the description of the situation which he uses, the conclusion which he wishes to draw. But the second is different. The Trojan case is based on independ-

ently describable facts (viz. those about IEM), facts which the Greek may, if he wishes, deny, but which are not simply conjured up by one's way of describing the situation.

Let me explain these unhelpfully abstract remarks. On p. 213 Evans issues a warning about 'a danger inherent in all reflection about self-consciousness'. He writes:

... we have been able to take the subject of thought, and his identity, for granted ... Now there is no harm in continuing this way of proceeding when we come to consider self-identification: indeed it is unavoidable. But we must realise what we are doing. We are building the subject's identity over time into our description of his situation. This may make it appear that he has an infallible knowledge of what is involved in this identity; but the appearance is nothing but an artefact of our way of describing the situation (ibid. pp. 213-4).

Evans' point (which will become clearer) is one of those brilliant aperçus with which The Varieties of Reference is well-endowed; but in examining its applications one must be careful to prevent one's admiration from quite overwhelming one's judgement.

Its first application is in connection with Anscombe's suspicions arising from the fact that there can be no 'unnoticed substitution' of the object of self-reference. Perhaps it is intelligible to suppose that I wake up one morning and find that, like the central character in Franz Kafka's Metamorphosis, I have changed into a beetle. But it is not conceivable that I have changed into a different person, such that I fail to identify, by means of 'I', the same object that I identified in like manner yesterday. In contrast, it is quite possible that an 'unnoticed substitution' may occur of the object I intend to refer to by e.g. 'that man' in our earlier example. This is logically possible even if I am looking at the poor fellow continuously for some period from the time at which he fell out of the carriage; a fortiori if the subject is not under continuous observation and I have to pick him out later in an identification parade. The Trojan case suggested by Anscombe is that the 'logical guarantee' of correct re-identification present in the case of 'I', but of no other

expression, is a symptom of non-reference. Unfortunately for Anscombe, the following two responses to her suggestion are correct:

- (i) the 'logical guarantee' of correct re-identification is empty, by Evans' aperçu, and so cannot indicate non-reference.
- (ii) further, there is a case for saying that if we describe the behaviour of the other indexicals in a relevantly similar way to that in terms of which we describe 'I', then we will obtain a similar (and similarly empty) guarantee of 'no unnoticed substitution' with them too.

The second point will be dealt with at length and shortly. With regard to the first point, Evans writes:

But, of course, the 'logical guarantee' is simply produced by Miss Anscombe's way of describing the situation, in terms of one and the same subject having thoughts at different times. It is a simple tautology that, if it is correct to describe the situation thus, the self-identifications are all identifications of the same self, and hence it cannot be a reason for anything ([1982] p. 214).

Now, there are indeed 'mind-stretching' puzzles about the self, and the fact that Evans has just delivered a knock-out blow does not detract from the fact that this issue between Anscombe and Evans constitutes one of them. It raises in particularly acute form a problem apparent throughout the course of the Homeric Opposition - that how you choose to say it seems to determine what you will say. Everything that one says about the self seems, in Evans' words, simply to be a product of one's way of describing the situation!

Before we consider this defeatist response in connection with the stronger Trojan case, it would be well to discuss whether there is indeed a disparity between 'I' and the other indexicals with regard to the possibility of unnoticed substitution. There will be such a disparity as long as we fail to regard constancy of subject (that is, of speaker or thinker) as just another aspect of the context of utterance or thought. It is this context which in the

case of indexicals alone, must be specified before we can know the referent. Let us consider the contextual rules for the different indexicals:

1. In order to determine what is the referent of 'I', one needs to know who is thinking or speaking.
2. In order to determine what is the referent of 'here', one needs to know where the thought is entertained or the utterance made.
3. In order to determine what is the referent of 'that \emptyset ', one needs to know the context which fixes the object of demonstration.

Now, Anscombe assumes tautologically that in re-identification in the case of 'I', it is the same person who is thinking or speaking. If we make analogous assumptions with the other indexicals, we will get the same empty guarantee of correct re-identification. That is, if we specify that the thought is entertained or the utterance made in the same place, 'here'; or that it is the same context which is fixing the object of demonstration, 'that man', say. So the disparity between the indexicals regarding 'no unnoticed substitution' is mere appearance.

If this view is correct, then not only is the 'no unnoticed substitution' of 'I' in fact empty, but a similarly empty feature can be generated for other indexicals - 'an artefact of one's way of describing the situation', to use Evans' words. The existence of a genuine such parity between the indexicals would be one less reason for sharing the Trojan suspicions about the referring role of the 1st-person pronoun, or for endorsing their claims about its uniquely non-referring role amongst the indexical expressions.

(ii) The unique IEM guarantee and the substantial Trojan case

But the second Trojan case is not, at first sight at any rate, any mere artefact of one's way of describing a subject's situation. This case, which I have already summarised, is not one about immun-

ity to error in re-identification (in the sense of guaranteed co-reference to oneself at different times), but is about the immunity to error through misidentification of certain past- and future-tense statements made on certain grounds. Not a point about the reference of certain terms (nor of identification in the sense of reference) but a point about certain statements and their immunity to a certain kind of error. This case is based on the disanalogy between 'I' and other indexicals brought out in (IEM₇). To re-iterate: It is not just that in IEM-utterances (as in all 'I'-utterances) I refer to myself, and that one such utterance and any past- or future-tense transformation of it will co-refer. That is the 'no unnoticed substitution' phenomenon, a reflection of the empty 'guaranteed reference'. Rather, it is that there will be an IEM-basis available for a past-tense, and often for a future-tense, transformation of a present-tense 1st-person IEM-judgement. Thus we have the idea of a persisting subject, who can at different times make the same statement about himself, which if he bases it on a certain type of ground, will not be falsifiable due to misidentification. Two examples will clarify this claim. First, to take the now-hackneyed example. I say, on the basis of my intention to walk across the Tay Bridge on a day which is not the calmest of summer, 'My hair will be blowing in the wind (I'd better get out the Brylcreem)'. But I then forget to apply the famous hair-product. So, I walk across the Tay Bridge, reflecting on the pleasant sensation of my hair blowing in the wind, and give verbal expression to the appropriate IEM-judgement. Later I say 'My hair was blowing in the wind', on the basis of my recollection of the sensation. This is the sense of guaranteed identification over time which the Trojan is after. The second example is a little different. This time I do apply the Brylcreem, and start walking across the bridge. Unfortunately a freak gust of wind picks me up and hurls me over the parapet - but I happen to land on the deck of a passing marmalade-carrier (Dundee being a leading centre for the production of this important foodstuff). I strike my head a glancing blow on the deck and, in a Baldy-type misadventure, suffer consequent amnesia. Throughout this extraordinary episode the Brylcreem has done its work and my hair never blows in the wind. However, to cut a long and tedious story short, I later come to believe on the basis of inaccurate testimony (a case of mistaken identity) that my hair had been blowing in the wind after all. Clearly, then, not any transformation of the

present-tense utterance will do; there has to be some continuity in the basis for the judgement.

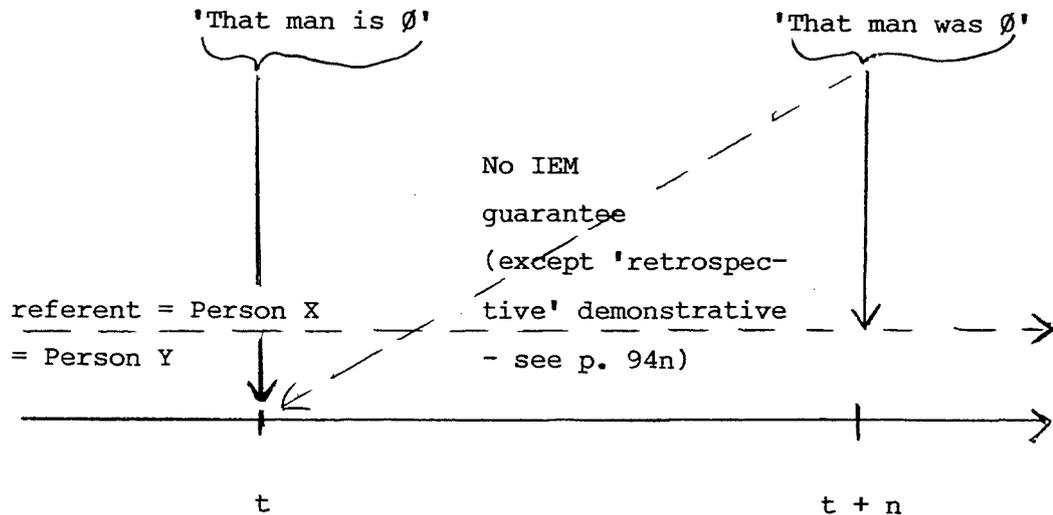
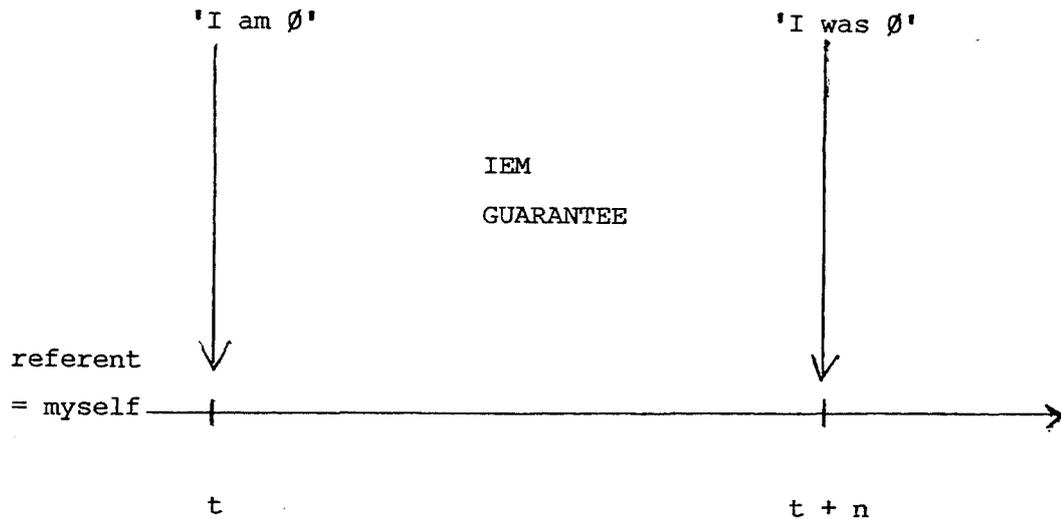
Now, that Evans' first application of his aperçu has no connection with IEM, and thus with the second Trojan case, is brought out if we consider that his criticism of Anscombe applies to non-IEM utterances. If one of the gentle giants in our earlier example (p. 11) says 'I weigh 25 stone', and later says 'No, the weighing machines was rigged [sic], I only weigh 24 stone', there is guaranteed re-identification in the second utterance of the object identified in the first, but neither utterance is IEM.

So it is the IEM of past- and future-tense transforms of present-tense utterances, and not this latter feature, which is captured in (IEM₆), and as noted earlier, it does not seem to be present in the case of 'that' (nor of other indexicals e.g. 'here'). We have reason, then, to doubt a comment which Evans makes on Wittgenstein's move from IEM to the 'no-reference' view:

This seems to be just a mistake. For we have seen that immunity to error through misidentification is a straightforward consequence of demonstrative identification ... [which] is, precisely, a way in which a thought can concern (be about) an object (ibid. p. 218).

(Though see p. 94n.) But before discussing arguments from IEM to 'no-reference', we need to get clear the difference in behaviour between 'I' on the one hand, and the demonstratives (and other indexicals) on the other. In so doing we will examine whether Evans' second application of his aperçu does indeed bear on the second Trojan case.

Two diagrams may help:



In the second diagram, X and Y may be the same person but the absence of a guarantee that they are is not, in the stronger case, the Trojan point. Rather, it is that sentences of the form 'That F is Ø', when transformed into the past or future tense, and uttered on the same grounds, can never exhibit IEM. There are no grounds such that an utterance of that sort will carry a guarantee of IEM - though of course many such utterances will in fact not exhibit error through misidentification. (A past- or future-tense demonstrative utterance may by itself carry the guarantee of immunity of this sort. 'That man was in a street-fight', said on the basis of his present appearance, is so immune - but IEM is not preserved over time, in the sense of formulation (IEM₇) for the reason given on

pp. 25-26).¹ In contrast, as already explained, every present-tense IEM-judgement in the 1st-person seems to imply the potential for a past-tense IEM-judgement to be made on the same grounds, and often to look back to a future-tense IEM-judgement that could have been made. (This is a complex matter though. It looks as if, for subsequent past-tense judgements, the grounds of the present-tense judgement are simply preserved in recollection. But take a future-tense IEM-judgement made on the basis of knowledge of one's intention; the present-tense transform of this could not be asserted on the basis of intention recollected.)

In his second application of his thought, does Evans quite grasp the distinction between the two Trojan cases? Does his argument fully meet the stronger Trojan case? Well, it is not clear whether Evans has really taken on board the features of IEM thus far outlined, and of which he subsequently gives a not dissimilar treatment to that of the present writer. He seems to be giving a 'Parfit-style' objection to the phenomenon of IEM, but in a footnote he writes:

My point is not to deny that there is such a thing as criterionless self-ascription of anticipated properties ... But I do want to deny that this is a matter of a logical guarantee of an identity assumption (*ibid.* p. 215n).

This latter denial is intelligible, but as we shall see, it is false. Adopting Evans' example, take the utterance 'Tomorrow my hair will be blowing in the wind', made on the grounds, inter

¹ This question of the 'retrospective demonstrative' (a demonstrative judgement in the past-tense which exhibits IEM) is a vexed one, in fact. The Trojan point, endorsed above, that there is no cross-temporal identification guarantee because a demonstrative thought is not repeatable (pp. 25-26 above), may cause anxiety. Is the special IEM of 'I' simply demonstrative IEM plus the empty 'no unnoticed substitution'? This thought may provide a short-cut to a Greek refutation at what in the light of the plausible arguments in section (2.8), is the weak link for Trojanism.

alia, of intention, and thus prima facie IEM. It is, in this case, not necessary, nor is it possible, to exercise criteria of subject-identity. Evans' contention is that this criterionless self-ascription of anticipated properties does not mean such ascription cannot be mistaken, that there is a logical guarantee of correctness in identification involved. But as we shall see, there is such a logical guarantee - IEM. Evans' concessionary footnote is thus inadequate.

(iii) Parfitian considerations¹

To return to the 'Parfit-style objection'. It can be presented in the following way:

Suppose someone who is an unthinking adherent of the Simple View of Personal Identity, makes the statement "20 years ago I fell out of a carriage". He makes this statement on the basis of q-memory.² But in fact, the self that suffered the accident is merely psychologically continuous with the speaker. So here is a non-IEM transformation of a judgement which, if made at the time of the accident, would have been IEM. Acquaintance with the true facts about Personal Identity, its limited extent and relative unimportance, will make us realise that the above kind of failure of past-tense IEM (and, analogously, future-tense IEM) is widespread. And how are we to know, for any particular past- or future-tense statement in the 1st-person, whether the guarantee of IEM holds or not?

¹ The following section was written before I had a chance to look at Parfit's [1984], but no substantial point seems to be altered by my having adverted solely to the earlier articles. In fact, no essential thread of the main argument is missed by turning now to p. 100.

² Q-memories are allegedly 'phenomenologically identical' to and caused in the same way as ordinary memories but originate in the experience of a person other than oneself. The 'Simple View' is that of Parfit's opponent as described in his [1971].

Evans makes what I assume is this same point, but related to future self-identification, in which a Parfit-style explanation is implicit:

[It may seem that] a subject [has] an infallible knowledge of what it is for a state of affairs to concern his own future ... We might say: ... when time 't' comes, [the subject] will know whether or not the hypothesis that he expressed earlier by 'I'll be in pain at t' was or was not correct; so all he has to envisage, when he envisages the future state of affairs of his being in pain, is a future pain ... It is not possible for the subject to have got hold of the wrong person at time 't'.

Evans continues:

... there is something correct about this [viz. that there is ... criterionless self-ascription of anticipated properties] ... Of course it is not possible for the subject to have got hold of the wrong person - as the case is described, there is a logical guarantee of adequacy. But this, again, is an artefact of our way of describing the situation: it certainly does not show that, just by envisaging future situations, a subject has a complete and clear conception of what it is for a future state of affairs to involve himself. The 'method of verification' has a presupposition ... that the subject who exists at t and 'remembers' the hypothesis expressed earlier is the person who made the hypothesis, and hence is the person whom it concerns. And this is something of which he can have no logical guarantee (ibid. pp. 214-5).

This is a difficult point to unravel. But first, Evans does not make it explicit that he is dealing with IEM. True, the example 'I'll be in pain at t' could be IEM; but it need not be. (Say I'm one of six prisoners, and know that at t one of us will be tortured, and suspect that, as the most likely informant, it will be myself. In this case, my grounds for believing someone will be in pain at t will survive the discovery that I am in fact to be released before time t and that it will be one of the others who will be in pain then.) In one sense, Evans writes, 'It is not possible for the subject [now] to have got hold of the wrong person at [future] time

t', in making his prediction - in using 'I', the subject is guaranteed a reference to himself. But Evans does not bring out the different sense in which an extra identificatory guarantee holds for some (viz. IEM) utterances and not others. For in the prisoner example, of course, it is not true that the subject, just by 'envisaging future situations ... has a complete and clear conception of what it is for a future state of affairs to involve himself'.

But when the subject seems to have such a conception, both for future and past states - that is, in the IEM cases - must we say, with Evans, that this apparent truth is a mere artefact of our way of describing the situation? And that there is a presupposition - that the subject who exists at t and 'remembers' the hypothesis is the person who made it, to use Evans' example? Is this really a presupposition?

Only if we take the Parfit-style objection (a) seriously and (b) treat it in a certain way. The objection seems to involve denying that the phenomenon of IEM exists; or rather, denying that one can know in any particular case, that the immunity obtains. This, on the grounds that one can identify, on the basis of q-memory, q-intention etc., certain past or future mental states of a subject in the same way that one identifies one's own, yet these states belong to a self merely psychologically continuous with oneself. Only if this is admitted as a possibility does it become a presupposition that the remembering subject is the hypothesising one; but in admitting this, the independently describable facts of IEM are being denied, and it is here that the dispute lies. The Trojan, in his second and stronger case, has not committed the Anscombe fallacy uncovered by Evans' aperçu.

I have said that the dispute is over whether the apparent facts of IEM do indeed hold, or whether the existence of q-memory etc. would limit these facts in an anti-Trojan way. But we need to refine the Parfitian claim before we assess the dispute. As I said above, we have to treat it in a certain way. Let us look back at (IEM₇), which referred to the making of the same 1st-person assertion at different times. But presumably 'I've just fallen out of the carriage' and 'I fell out of the carriage' uttered 20 years

later by a merely psychologically continuous self, are not the same assertion. Sameness of reference is required, but in fact the subjects are not identical. Since, the Trojan argument continues, the latter assertion is not a genuine transformation of the former, we should not expect it to be IEM. The IEM over time we are concerned with relates only to transformations of present-tense IEM-utterances. With the extinction of the self that fell out of the carriage (where 'self' simply means 'person!'), the original assertion is no longer available to be made in the 1st-person (though of course it can be made in the 3rd-person; unless we are attracted by the solipsistic strain of Trojanism).

This Trojan defence is effective only if we say that a competent 'I'-user cannot be mistaken about the duration of his present self, the Parfitian objector will rejoin. However, the speaker who uses 'I' in talking of a merely psychologically continuous self is not using it uncomprehendingly, but simply mistakenly. And the consequence of this is that a comprehending 'I'-user may not know whether some past-tense IEM utterance in the 1st-person is a genuine transformation of some present-tense IEM utterance. The Trojan wants to say that, if selves are less durable than bodies, IEM can still apply throughout the shorter duration of a self. But because the subject can be mistaken about the status of a 'former' or 'future self' (to use Parfit's façon de parler), judgements in the 1st-person can be mistaken owing to an error in misidentification.

It is becoming apparent that the disputants of the Trojan plain, and perhaps the indulgent reader too, are wearying of combat, but I will suggest two lines of Trojan response to this last point that seem to render it ineffective as a Greek objection:

- (i) The error that the subject in the example makes is one of reference, not identification - or, if we prefer, given the ambiguity of the latter term, we should say that the statement does not fail to exhibit IEM, but rather contains a term which fails to pick out its semantic referent. If we recall, it was stressed in our earlier account of the distinction between IEM and guaranteed reference, that both IEM and non-IEM 1st-person utterances refer to the speaker.

The present aberration is thus not connected with IEM. (The Trojan would need to show why this kind of response cannot be deployed, in an anti-Trojan way, in the case of cross-temporal demonstrative identification discussed on p. 94n.)

- (ii) We don't have the q-concepts which are integral to Parfit's account of the limits of Personal Identity. But even if we did, would we have to concede that, say, the kind of thing known on the basis of other-originating q-memory would be indistinguishable from the kind of thing known on the basis of a member of the sub-class of q-memories, 'genuine' memory? I might know of someone else's intention in the way I know of my own, but does that mean I might be mistaken about whether the intention was someone else's or my own?

I am thus unconvinced that the Parfit-style objection (if Evans in fact endorses it) offers a serious Greek objection to the account of IEM from which the Trojan seeks advantage.

2.8) Trojan Arguments from IEM (II): The Wittgensteinian Route to the 'No-Reference' View

(i) Two Trojan arguments

What the phenomenon on which the Trojan arguments are based is, and what it is not, should be clear. So I will now spell out these arguments in detail, adverting first of all to their origin in that passage of the Blue Book previously mentioned (on p. 3). For whilst the passages in the Philosophical Remarks constitute not much more than an illustration of Trojanism, that in the Blue Book provides the basis for a justification of it. After the aforementioned discussion of 'I'-as-object and 'I'-as-subject uses, Wittgenstein writes:

... it is as impossible that in making the statement "I have toothache" I should have mistaken another person for myself, as it is to moan with pain by mistake, having mistaken someone else for me. To say "I have pain" is no more a statement about a particular person than moaning is. "But surely the word 'I' in the mouth of a man refers to the man who says it; it points to himself; and very often a man who says it actually points to himself with his finger". But it was quite superfluous to point to himself. He might just as well only have raised his hand.

The example of pain and moaning does not help to dispel the confusion between IEM and avowal mentioned previously (pp. 18-21); nor is Wittgenstein exactly expansive on the connection between IEM and the fact that 'I' is non-referring! Some compensating expansiveness on the part of the present writer is therefore called for.

The arguments from IEM to 'no-reference' are alluded to rather than stated in the post-Wittgenstein literature, but it is possible to present two forms:

- (i) There are no criteria for, and thus no possibility of justification of, any identification of a subject in 1st-

person IEM-judgements. Hence these judgements cannot involve an identification of (and thus reference to) a subject.

- (ii) One cannot fail to identify correctly the subject of 1st-person IEM-judgements. But then there is no sense to the idea of succeeding in identifying (and hence referring to) a subject either.

There are analogous arguments from the criterionless and incorrigible self-ascription of avowable states to the non-assertoric thesis of avowals. This is hardly surprising in view of the structural similarities of IEM-judgements and avowals (viz. that avowals exhibit the IEM phenomenon plus a similar phenomenon applying to the predication component - if one may so put it (see p. 23)). There is this difference between the conclusions of the two sets of arguments (the 'no-reference' and the non-assertoric theses), however. In the case of the non-assertoric thesis we have very strong additional reasons which urge its acceptance (viz. the flight from Cartesianism), but in the case of the 'no-reference' thesis, there are just the arguments. It is this absence of a compelling motive which makes the 'no-reference' view appear recherché and idiosyncratic, whereas in contrast the non-assertoric thesis, though commonly albeit mistakenly viewed as a desperate expedient, is quite intelligible.

The first argument is less powerful. The phenomenon it refers to worries Strawson, however, who says it 'lies at the root of the Cartesian illusion' - and therefore, one may add, at the root of the other 'dualistic' illusion, Trojanism (see his [1959] p. 98). His positive account can be seen as directed against both 'illusions':

'I' can be used without criteria of subject-identity and yet refer to a subject because even in such a use, the links with criteria are not in practice severed ([1966] p. 165).

The links according to Strawson are: that the utterer of the 1st-person sentence is one for whom there are empirical criteria of personal identity, and that he is one who would recognise such criteria to be applicable to himself. Well, it is rather lame of Strawson to plead this! Let's say a bit about the links and the

lameness. What Strawson seems to be driving at is something like the point about truth-value links already deprecated in the discussion of arguments from the Generality Constraint (pp. 49-50). For instance, say Strawson says, 'I am perambulating around Oxford'. There are no 'criteria of subject-identity' here (the judgement is IEM¹). But there are truth-value links between this sentence and: 'Strawson is perambulating around Oxford' (for instance). And in the latter case there clearly are empirical criteria of subject-identity. Indeed, for any 1st-person utterance Strawson or anyone else makes, there are truth-value links with 3rd-person utterances exhibiting criteria of subject-identity. However, if one regards this as a telling point against Trojanism, then that doctrine gets a rather short run for its money. As has already been urged, though, co-reference is simply the most economical explanation of these links, and is not a datum. Hence, the Trojan may (and will) concede the applicability of such empirical criteria and their links with the 1st-person, without conceding that such links establish co-reference of 1st- and 3rd-personal expressions and the identity of self and person. (He may give an account in terms of verification - see Anscombe op. cit. p. 61.) It is not my intention to pursue the first argument further, however, since the second argument is more compelling.

(ii) 'If you can't be wrong, you can't be right either'

Graham Nerlich [1967] has attempted to characterise the second form of argument, which relies on the slogan 'if you can't be wrong, then you can't be right either'. He says arguments of this type have been common in philosophy since the publication of the Philosophical Investigations,² and cites the example of

¹ And it is IEM that Strawson is referring to - even though he thinks it is limited to certain kinds of psychological self-ascription.

² The slogan itself seems to underlie the discussion of sensation 'S' in the course of the Private Language Argument (PI 258-70), and a closely related principle is clearly implied in the course of the later discussion of intention:

D.M. Armstrong, who writes 'We can speak of gaining knowledge only in cases where it makes sense to speak of thinking wrongly that we have gained knowledge' [1963]. Nerlich tries to elucidate the principle behind the argument, finds that it licenses absurd results, and offers a diagnosis: the principle involves a fallacy in which there is the shift of a modal operator. His version of the principle is the following:

'Fx' is meaningfully applicable to cases where 'Gx' is true
 \equiv possibly $(\exists y)(Gy \sim Fy)$ (ibid. p. 303).

He writes that 'Armstrong has agreed, in conversation, that it fairly represents the principle he invoked'. It is clear that Nerlich did not show Armstrong the rest of the article, however; otherwise this doughty philosopher would surely not have made such a foolish admission.

The principle is meant to underlie a form of argument employed by Trojanism. So it will come as no surprise that the non-cognitive thesis of avowals will, so Nerlich would contend, appeal to such a principle. And we can see by substituting 'x knows that he is in pain' for 'Fx', and 'x is in pain' for 'Gx', and similarly for 'Fy' and 'Gy', that the principle does have the thesis as an implication. Unfortunately for Armstrong & co., apparently, it also has the implication that we cannot say of any unmarried adult male that he is a bachelor (substituting 'x is a bachelor' for 'Fx', 'x is an unmarried adult male' for 'Gx' and so on). But of course no such principle can underlie the intuitively appealing form of argument in question. The reason for this is clear - the principle underlying the argument must make reference to what is required for the subject not just to possess some property, but to be credited with some achievement. The kind of achievement under consideration in the avowal and IEM cases is a cognitive one, and suggests the following principle:

'But in the sense in which I cannot fail to will, I cannot try to will either' (PI 618).

- (K) 'X knows that p' is meaningfully assertable only if possibly
 (p & X believes that $\sim p$)¹

This principle would rule out as examples of knowledge the following knowledge-ascriptions:

- (i) X knows that it is he himself ('I'-reflexive) who is that person whose hair is blowing in the wind (where X knows 'My hair is blowing in the wind' on an IEM-basis).
- (ii) X knows that he is in pain.

In the first case, clearly X can know 'My hair is blowing in the wind' since he may be wrong; but there can be no error in the identification component and it is this latter which cannot be a genuine item of knowledge for him, the principle implies.

The motivation behind (K) might be, as has been implied, the idea that, in cognitive as well as other kinds of achievement, the possibility of failure is required for the notion of achievement to be applicable. A separate motivation arises from a suspicion pertaining to knowledge itself; a suspicion that apparent truths the obtaining of which is conceptually connected with their apprehension by a knowing subject are not genuine items of knowledge (or belief). The conceptual connection is such that the obtaining of the truth 'Hamilton is in pain' is invariably concurrent with its apprehension by the subject: a remarkable fact! Perhaps this motivation has no force independent of (or is just an unhelpfully abstract way of putting) the more specific anti-Cartesian motivation, pertaining to the particular range of apparent knowledge-claims involving avowable-states: that there can be no such knowledge-claims because of the Cartesian incoherences they would imply.

¹ Nerlich does consider incorporating 'belief' in his principle, but produces an unsatisfactory formulation and so rejects the idea. A principle analogous to (K), but dealing with assertion, is discussed below in connection with the non-assertoric thesis of avowals (p. 180).

The fact that the identification component of an IEM-judgement cannot properly be said to be known does not of itself establish that the subject-term is non-referential, of course. In any case, it is the immunity to error over time that ostensibly marks out the 1st-personal non-subject-referring IEM-judgements from the rest. So the principle must rule out as a cognitive achievement only a certain kind of 'success' in identification - otherwise of course demonstrative identification too would be under attack. But also its employment in any Trojan argument must include a corollary to the effect that reference always involves identification.

(iii) A first Greek response: identification can be 'passive'

The refinement above will not be pursued directly here. Rather, consideration will be given to how the Greek need not be committed to direct refutation of the IEM arguments, since he can deny their implicit assumptions. These assumptions are, first (as already mentioned), that reference always involves identification (or that the two are essentially equivalent¹). And second, that identification is essentially something active or intentional; it can never be passive or non-intentional. One can never identify an object without, in some sense, intending to identify it (though the ascription of intention here may well be problematic in the way that ascription of intention to follow linguistic rules is problematic - viz. because the level of description of behaviour is too basic). The Greek may deny the first assumption whilst agreeing with the second; or vice versa. Gareth Evans takes the latter course. He agrees with the first assumption, though not in the way the Trojan envisages, and attempts to distinguish two senses of 'identification', in only the first of which is it 'active'. His remarks occur in the context of his criticism, considered unsatisfactory above, of Wittgenstein's move from IEM to 'no-reference' (see p. 92):

In one sense, anyone who thinks about an object identifies that

¹ It might be said that there is an argument for this view in Evans' deployment of 'Russell's Principle' - that reference requires identifying knowledge. But read on.

object (in thought): this is the sense involved in the use I have just made of the phrase 'demonstrative identification'. It is quite another matter ... for the thought to involve an identification component - for the thought to be identification-dependent. There is a danger of moving from the fact that there is no identification in the latter sense (that no criteria of recognition are brought to bear, and so forth) to the conclusion that there is no identification in the former sense ... (ibid. p. 218).

Evans is claiming that, far from reference essentially involving the active components of identification, and failing to occur when these are absent (as the Trojan wants to claim), rather there is a sense of identification (the passive sense) in which it is equivalent to reference. At least, that is how his remarks may be construed if one takes object-direction in thought as the correlate of reference. Two points must be borne in mind in assessing this 'passive identification' response to the Trojan arguments:

- (i) 'Active identification' and 'passive identification' are just labels for concepts as yet unclear. In particular, we would need to be told more by Evans of what his 'identification without criteria etc.' consists in.
- (ii) Evans' comments may in fact turn out to be decisive against the first argument from IEM (regarding criteria), yet not the second. Is thinking about an object something that 'just happens' to me? If not, and if it therefore makes sense to intend to think of an object and yet fail to do so, the familiar strictures apply in the use of identification qua object-direction too.

There is in the Greek annals another attempt to show that the principle behind the second argument from IEM need not yield a Trojan conclusion. Shoemaker, in his seminal [1968], takes the alternative course to Gareth Evans, denying the first Trojan assumption that reference always involves identification - though the upshot is the same (viz. reference without ('active') identification). He begins by endorsing the idea that 'identification necessarily goes together with the possibility of misidentificat-

ion'. Yet he also concedes (as he must) that there are predicates 'the self-ascription of which is immune to error through misidentification' (p. 562 - indeed he originated the phrase, though without noting that it is certain uses of predicates which are so immune). He writes:

I see nothing wrong with describing the self-ascription of such predicates as manifestation of self-knowledge or self-awareness ... the main source of the [Trojan] trouble here is a tendency to think of awareness as a kind of perception ... [but] self-awareness does not involve any sort of perception of one's self, i.e. does not involve what I have called 'being presented to oneself as an object' ... the way out of [the] incoherence is to abandon completely ... the perceptual model of self-knowledge (ibid. p. 563-4).

Shoemaker goes on to argue that the existence of a class of IEM-judgements is necessary for self-ascription to be possible at all:

It is possible for there to be self-ascription involving self-identification only if there are some self-ascriptions (the IEM ones) which do not involve self-identification.

The case he makes need not concern us here. What is of interest is the way Shoemaker anticipates Evans in proposing a variety of self-reference (or, as he puts it, self-awareness / self-knowledge) without self-identification in the 'active' sense. But as with Evans' comments, we need to be told more about what this non-perceptual 'non-identificatory' self-knowledge is like. On this question Shoemaker is silent.

(iv) A second Greek response: 'Active' identification can be infallible

I will not pursue further the option of 'passive' identification, but return instead to the unresolved question of how the Greek can give an account of an 'active' identification that can never fail. Consider two kinds of case. First, I run a number of what seem to be races, and win them all. Second, I make a number of what

appear to be self-identifications, and make them all correctly. What are the explanations for these occurrences? In the first, it turns out that I am the only 'competitor' - so I haven't in fact 'won' anything. What explanation of the identification case will prevent us saying something analogous? Well, most IEM non-avowals are grounded in identification of an avowable-state which is an immediate sensation. But the concept of sensation excludes any error in identification when ascribed on an immediate, transparently non-inferential basis rather than on the basis of behaviour - i.e. when self- rather than other-ascribed. (Recall Mrs. Gradgrind's final mistake!) The explanation of why most IEM-judgements have the property they do, then, is that they require the exercise of a distinct ability to identify sensations, which ability, in the kind of case under consideration, necessarily involves identification of the owner. This is a substantial explanation (or justification), unlike the simple grammatical explanation in the one-man race which pointed out the connection between winning and competing. It is a genuine explanation of success in identification which characterises the class of IEM-judgements independently of the property of IEM, in terms of their grounding in immediate sensation.

Three points arise:

- (i) The account needs to be supplemented to cover IEM-judgements not grounded in sensation (e.g. expressions of desire or intention).
- (ii) The account is a 'grammatical' one nonetheless; it does nothing to establish that the continued success in identification is contingent (and how could it?). So the suspicious logical guarantee of correctness remains, the Trojan can urge.
- (iii) The Trojan may have a rival story to tell.

I will first say a little about the third point. For the Trojan may well rejoin: 'Isn't the explanation of the "correct identification" simply that the concept of sensation, in the range of cases under consideration, makes no provision for the identification of an owner?' Well, this is what we would expect the Trojan to say (and

the fact that he can do so brings to mind the Evans-inspired pessimism mentioned on p. 89!). It is simply an illustration of his viewpoint - an alternative picture to that of the Greek. As such it requires supplementation with the sort of argument indicated by point (ii) - a circularity argument, in effect. Just what circularity should the Trojan point to in the Greek account? He might try to urge that the Greek is trying to explain incorrigible identification of the subject in terms of incorrigible identification of sensation - the very same problematic kind of phenomenon. But he would perhaps be ill-advised to try this move, for reasons to be discussed shortly. A better alternative would be to urge that the Greek is explaining IEM as it appears in the entire range of IEM-judgements in terms of the IEM of avowals - since this is the 'cash-value' of the aspect of sensation dwelt on. The Greek, to counter this latter charge, would need to maintain that the feature of sensation he is pointing to (the necessary identification of an owner) is not just the IEM phenomenon as it applies to avowals (though it can be presented that way - and has been in our neutral characterisation, pp. 23-24 above). It is, rather, a basic, unexplained phenomenon, which itself explains IEM.

I will not say anything further on this dispute, nor indeed on the arguments from IEM. The arguments remain attractive, their consequences - in terms of the disunity introduced into our account of 'I' - less so. The inconclusiveness of this state of affairs should not be too worrying however - since Trojanism as thus far considered (i.e. the 'no-reference' view of 'I') is an unstable position. Recall the charge which it was claimed above that the Trojan might be ill-advised to make - viz. that the Greek is trying to explain incorrigible identification of the subject in terms of incorrigible identification of sensation. Well, to this there is the following Greek riposte: 'You claim to find difficulty in the notion of incorrigible self-identification. Very well, then I pose the following dilemma. Either you must accept that analogous arguments apply in the case of incorrigible identification of sensation - in which case you are committed to the non-assertoric thesis of avowals (which is, of course, anathema). Or else you must drop them in the case of self-identification. Your current position is therefore quite unstable'.

It does seem that the riposte is justified - and indeed that the 'no-reference' view of 'I' is unstable, though not perhaps in the sense that the Greek portrays it. What the riposte establishes is that Trojanism must at least include a non-assertoric account of avowals. This is because it is hard to see how, with neither the identification nor the predication component referring, an avowal could be an assertion. (Though the converse would not hold - that an avowal is not an assertion does not imply its components are non-referential.) It is difficult, then, to see how the 'no-reference' view can be assessed in isolation then - we will need to move on to discuss the non-assertoric account of avowals. However, the fact that the Trojan case thus far remains 'not proven', combined with the non-Trojan outcome of the later discussion of avowals, justifies the Greek stance of the chapter in between - on Self-Consciousness.

CHAPTER

THREE

It is the task of this chapter to elucidate the self-reference of 'I'. The bipartisan policy of Chapter 1 is now clearly abandoned, and we are adopting the Greek perspective; but all the phenomena to be discussed would need to be acknowledged by any Trojan diehard. (The resolution this implies pertains only to the first part of the Homeric Opposition - the validity of a 'Trojanism of avowals' has yet to be decided, of course.) An understanding of the self-reference of 'I' is clearly fundamental to an understanding of self-consciousness; but it is only the first step. It would be an impossible task to deal with the general topic of self-consciousness within the one chapter which is all that can be allotted. Beyond self-reference, some consideration will be given to the role of IEM in the constitution of self-consciousness, but more than that will not be attempted. The question of the relation of self-consciousness to 'consciousness' or sentience, and of the range of capacities required for the former, is left almost wholly untouched. Clearly, however, the discussion of avowal in Chapter 4 must have a major bearing on these questions.

Any account of the self-reference of 'I' must begin with the feature indicated by Anscombe and already discussed in Chapter 2 - the feature missing in the fortuitous self-reference of Oedipus and Emma. The self-reference effected by 'I' is one in which the speaker knows and intends that he refers to himself ('I'-reflexive). When this circularity-inducing requirement is satisfied, let us talk (following Nozick, pp. 113-14 below) of reflexive self-reference (as in 'indirect reflexive', 'token-reflexive', etc.).

How should one try to spell out what 'reflexive self-reference' entails? The chapter begins with Nozick's account. Quite lengthy but justifiably impatient consideration is given to this depressing account, and then we move on to the more fruitful ground of Anscombe's discussion. Here, the points made in Chapter 2 are taken up and developed further. The problem of circularity is treated and the self-reference principle of Evans is discussed. Finally, there is discussion of Anscombe's 'A'-users, whose self-reference is allegedly not self-conscious.

3.1) Nozick on 'Reflexive Self-Reference'

(i) 'Self-reference' and 'auto-reference'

Nozick in his [1981] attempts to explain reflexive self-reference and thereby reflexive self-knowledge (pp. 71-114). This account and (so I am told - I could not manage any more) the rest of the book exhibit the author's characteristic self-indulgence to a grotesque degree - sloppy, careless, little evidence of editing, at times incoherent, often tasteless - the whole being informed by the peculiar idea that the most interesting facts are the facts about R. Nozick himself. Nevertheless it is probably worth enduring the resultant battering of one's sensibilities in order to seize upon the author's one undoubted equivocation and upon the few grainy pearls deep-buried in the bloated oystery mass.

The equivocation comes at the beginning and it is a serious one from which the account does not really recover. My present purpose is therefore really diagnostic; also the recovery of accidental insight. Nozick wants to explain the difference between self-conscious or 'reflexive' self-reference and fortuitous self-reference as in our 'Emma' example (p. 58) or in Oedipus' search for 'the person whose acts have brought trouble to Thebes'. He writes:

For a person X to reflexively self-refer is not merely for X to use a term that actually refers to X; this omits as internal to the act of referring that it is himself to which he refers (ibid. p. 72).

This is of course familiar from our discussion of Anscombe (pp. 58-62); but note that it is the person who is self-referring. Nozick seems to forget this fact when he goes on to postulate a class of reflexively self-referring terms :

Some indexical terms have a reference that not only varies with the context of their utterance, but also depends essentially on the very utterance in which they appear: for example, "this

very phrase" refers to that phrase itself, and "I ..." refers to the producer of that token itself. Let us call such linguistic devices reflexively self-referring.

I will turn shortly to the idea that, equivocation apart, there may be something in what Nozick writes; but first let us conclusively establish that there is an equivocation. He attempts to explicate 'reflexive self-reference' first in terms of the joint requirement of rigidity of reference (as used in the actual world, the expression refers in any possible world to that very same thing it actually refers to in the actual world), and (in virtue of the term's sense) necessary self-reference. This is inadequate, Nozick contends, because of the counter-example of self-reference via Godel numbering (ibid. p. 75), so he goes on to consider the more promising idea of 'reference from inside', of which more anon. Here he is clearly talking of self-referring terms :

With reflexive self-reference, it follows from - is part of - the sense that the term necessarily self-refers in virtue of a feature bestowed in the token act of referring (ibid. p. 76).

Yet shortly afterwards, Nozick asks

... to explain how 'I' refers, must we use the notion of reflexive self-reference, speaking of someone producing a token with the intention of reflexively self-referring?

In this case it is the speaker who is self-referring. Let us call the phenomenon of the self-referring expression, 'auto-reference', reserving 'self-reference' for 'I' etc. Since it seems to preclude the possibility of determinate truth-conditions, auto-reference is a worrying thing; but it is not yet apparent that 1st-person utterances are infected with it - and in any case, we must first be aware of the distinction between self- and auto-reference.

Is there anything in Nozick's account thus far, despite the equivocation? Is there nevertheless something in common between 'I' and 'this very phrase' which illuminates the idea of reflexive self-reference? Regrettably, not as Nozick explains it there isn't. To return to his original explanation - 'I' is meant to have a refer-

ence that varies not only with the context of utterance, but which also depends essentially on the very utterance in which it appears (as does that of 'this very phrase'). Well, the context is: who the speaker is. Just as the context for 'here' is: where the speaker is. In the case of 'I', there does not seem to be any further 'essential dependence' of the reference on the very utterance in which 'I' appears. Insofar as there is such a dependence, it is simply a dependence on: who makes the utterance.¹

'But I've shown the difference between the reflexive self-reference of "I" and ordinary indexical reference by means of my brilliant notion of "reference from the inside", haven't I?' pleads Nozick. 'You haven't said much about that yet, have you!' Insofar as one can make sense of this notion, it does not, unfortunately, seem to mark out a distinction between 'I' and other indexicals. 'Outside references', we are told

...pick out their referent in virtue of some property or feature it independently has, a feature it has independently of being referred to then. Whereas, reference from the inside ... is peculiarly internal to the act of referring since that act refers in virtue of a feature created by or produced in that very act itself.

Nozick continues:

The act of referring is sufficient to bestow [that feature] but may the referent also have it independently, apart from the act of referring? (ibid. pp. 75-76)

Since the feature is, presumably, 'being the producer of the 1st-person utterance', the question looks unintelligible; but be that as it may, Nozick fortunately declares that 'We need not decide this now'. It is difficult not to become impatient in the face of such high-powered garbage, but it should be clear that, using the given

¹ Specifying the context in this way showed that there is no disparity between the indexicals in respect of 'no unnoticed substitution' either (pp. 89-90).

account, 'here', 'now' etc. will also 'refer from the inside'. If I say, standing outside Nozick's office at Harvard, 'There's a great philosopher working here', I will be referring to the vicinity of Nozick's office in virtue of a feature bestowed on it by the very act of reference itself - being the place where the 'here'-utterance is made. If I go on to say, uncharitably, and sotto voce, 'Actually he's not so good now', I will be referring to the time (around 1986, say) in virtue of a feature bestowed on it by the very act of reference itself - being the occasion when the utterance is made. (If I am indeed 'referring' to it at all.)

These considerations point to the fact that what makes 'I' different from the other indexicals like 'here' is not some feature it has in common with 'this very phrase', but a feature it has in common with personal proper names when used by their referents - viz. self-reference in the sense of reference to oneself. When it is realised that the alleged 'dependence on the very utterance' is not a shared feature of 'I' and 'this very phrase', the need for Nozick's subsequent account of the 'I' 'synthesised in the very act of reflexive self-reference' is undermined. The 'performative' aspect of 'I' is something shared with 'here' - but it is absurd to talk of 'place-synthesis in the very act of reflexive place-reference' (see below - p. 120).

This is not to say that the metaphor 'from the inside' may not be suggestive. It might be proposed that to use 'I' is to engage in an act of 'performative self-reference'. And there is a line of argument to Nozick's conclusion which, though it rests on several questionable assumptions, may be worth representing - as an indication of the kind of work a supporter of Nozick would need to do. It goes as follows:

- (1) 'I' and 'The producer of this very token' are (in some sense) equivalent in meaning.

So (2) (T_1) 'I am immobile' may be rendered as (T_2) 'The producer of this very token is immobile'.

- (3) In (T_2) (a) 'The producer of this very token' refers to the speaker (self-reference); (b) 'This very token' refers

to (T₂) (auto-reference).

(4) The reference in (a) is achieved by dint of the reference in (b).

(5) So is the reference in (T₁) (somehow).

It will be difficult to make out (4) and (5), and I do not propose to try here; it should further be noted that Nozick himself would probably be unhappy with (1) (see p. 118 below).

(ii) A muddle over IEM and 'reflexive access'

In the light of the vitiating equivocation at the outset, it is embarrassing to have to retail the seriousness with which the subsequent self-synthesising tour de force has been taken. W. Richards gives a favourable account of it in his [1984] ('admirably lucid'(!)), whilst Richard Rorty speaks of '... an intriguing quasi-Fichtean theory about the self as "self-synthesiser"¹'. Actually the account is intriguing, but not in the sense in which Rorty means it. Our present purpose is again diagnostic. What is going on behind this bizarre piece of philosophising?

In fact, between his equivocatory account and the self-synthesising tour de force, Nozick does raise an interesting question, though he makes a muddle of it. 'Self-synthesis' is a product of the equivocation and the muddle. The question is this: How is reflexive self-knowledge possible? What is the status of the self-knowledge expressed in reflexive self-reference? Nozick approaches this by asking 'How do I know I am referring to myself (when I reflexively self-refer)?'

Any adequate explanation of this knowledge appears, according to Nozick, to be circular. If I say 'I am in Cambridge', my knowledge that it is I who is referred to cannot be captured by substituting some descriptive expression for 'I'. It would always

¹ Article in 'London Review of Books' (20/5/82).

be possible for the subject to fail to realise his identity with the referent of that description:

Is my knowledge, "I am in Cambridge", captured by my knowing ... "this very producer of this very token is in Cambridge"? Yet this knowledge cannot consist in my knowing that the producer of the first part of that token referred to himself, for that involves the same problem of no guarantee that I (reflexively) know that I produced it ... We need to add the very phenomenon to be understood: his knowing that he himself produced the token (ibid. pp. 80-81).

This problem of Nozick's is, as far as I can tell (he rambles), the very same as Anscombe's problem of how to characterise the meaning of 'I' in terms employing the indirect reflexive (see pp. 58-62). (Though it is not the only problem to which Nozick seems to think self-synthesis offers the solution - see p. 122.)

The problem provokes from Nozick a shoddy collection of half-baked thoughts which shows that he has not really understood the post-Shoemaker literature on IEM. He concludes, from the insufficiency of anything but a circular account of reflexive self-reference, that 'we must each have a kind of access to ourselves which is not via a term or referring expression, not via knowing that a term holds true (of something or other)' (p. 81). Well, 'I' itself may surely be such a 'referring expression'; but what Nozick seems to be groping for (mistakenly, as it will turn out) is the phenomenon of IEM. Later, he refers to IEM explicitly and says that it is incompatible with a 'pre-existing I' (ibid. p.90). He says that there may be some way of observing oneself which cannot be used to observe anything else:

On this view, I know it is I who is in pain, by observing in that particular way that someone is in pain (And concluding that it must be I?).

But this does not fit reflexive self-reference that is non-psychological:

Moreover, my knowing that someone or other observes himself in

pain in this way does not constitute my knowing that I am the one who so observed himself. Reflexive access to ourselves, thus, cannot be a special mode of relating to ourselves as objects (ibid. p. 81).

It is doubtful that any self-respecting adherent of the Greek view will be troubled by these arguments. It can hardly be denied that there is a mode of self-observation which is sui generis. But clearly, when, by means of it, I 'observe' that someone is in pain (if that is not a tendentious formulation), I thereby know that I am that someone (I do not need to 'conclude' this) (see (IEM⁴), p. 8). Of course, my knowing that someone observes himself in that special way, does not constitute my knowing that I am that one (as Nozick puts it in the last passage just quoted). But my 'observing' someone in that way does involve knowing that I am that one. And to the Greek, of course, 'reflexive access' just is a mode of relating to ourselves as objects.

But we haven't teased out Nozick's puzzle. What of the non-IEM reflexive self-references, e.g. 'I was born in Brooklyn'? The reflexive self-reference there cannot be explained in terms of IEM. May I infer this judgement from IEM-judgements, though? But to talk in this way is to go on running together the ideas of IEM and reflexive self-reference as Nozick's careless account has been suggesting. Really, all the talk of special modes of self-observation which are IEM is beside the point here. If I know that I 'observe' someone in a way that is IEM, I know that it is myself that I 'observe'. But if I know that someone observes himself in an IEM way, clearly I do not know simply on that basis that it is myself who is observed. Nozick's original point is correct; reflexive self-reference cannot be explained without recourse to 'I' - there is no description under which I am guaranteed to recognise the identity of myself with the referent except 'I'. And this has nothing to do with IEM. This fact Nozick seems not to appreciate - both in his remarks just discussed and in his later positive account of self-synthesis.¹

¹ Despite noting that there are reflexively self-referring judgements which are non-IEM, Nozick later treats his self-

I said before that the self-synthetic tour de force is a product of the equivocation and a later muddle. The equivocation ensures that Nozick fails to perceive that the consequences he draws from the behaviour of 'I' could equally be drawn from the behaviour of 'here, 'now' etc. 'Self-synthesis in the very act of reflexive self-reference' means synthesis of a self in that very act. Why not talk of 'Place-synthesis in the very act of reflexive place-reference', then? The reason is that Nozick's account has a Cartesian plausibility - which place-synthesis doesn't! If one is in a Cartesian frame of mind, it might seem that his account offers an explanation of the knowledge-gap between 'I am F' and 'The \emptyset is F'¹. But in fact self-synthesis only seems like an explanation of this phenomenon if one muddles up reflexive self-reference with the expression of a mode of self-knowledge. There is, in fact, no one special mode of access of oneself characteristic of reflexive self-knowledge - rather, there are several different modes some of which generate IEM and some of which don't, but in all of which the Oedipus mistake is not possible. A fortiori, one cannot conclude, as Nozick does, that not only do we have a special 'reflexive access to ourselves', but that, moreover, it is not 'a special mode of relating to ourselves as objects' (ibid. p. 81).

(iii) The self-synthetic 'tour de force'

The over-emphasis on the performative aspect of reflexive self-reference, and the muddle over reflexive self-knowledge as a mode of self-knowledge which is IEM, lead Nozick to present his self-synthetic 'solution' to the problem of reflexive self-knowledge. He does this under duress - driven to 'entertain ... (but not quite to endorsing it yet)', as the only solution to the 'intractable problem'. He has already rejected two other solutions. One of these is even crazier than self-synthesis; it is the idea that 'the self places itself into its reflexive self-referrings'. This seems to be the idea that 'I' is identical with a self - which

synthetic account as an explanation of IEM - partially misconstrued as guaranteed reference of 'I' (see below p. 122).

¹ See above, pp. 117-8.

itself takes the place of e.g. 'Hamilton' in 'Hamilton is tired' when said by me:

By putting itself into the blank, mentally stepping forward into the space ('--- is tired') (as in the dance where "you put your whole self in") the self, we are imagining, succeeds in referring to itself (ibid. p. 83).

Such an account calls for psychiatric rather than philosophical diagnosis; and as such is beyond the competence of the present writer.

The other rejected account is more attractive, however; it is that reflexive self-knowledge is a 'basic phenomenon'. Shorn of its Nozickian excrescences, this amounts to the view discussed on pp. 132-34 that there is a non-vicious circularity in the explanation of the indirect reflexive from which 'I' is ineliminable. That is, self-consciousness must be manifested in and constituted by the use of a linguistic device with the features of the 1st-person.

This view acquires greater plausibility when we examine the rival substantial account involving self-synthesis. Nozick begins with the act of reflexive self-reference, characterised as 'this very act of reflexive self-reference'. In attempting to 'draw boundaries' round this act, we may be tempted to search for a 'pre-existing entity, the doer' - if we are attracted to the 'agent theory' that every act has a doer. (One may question whether this is a theory, but never mind.) On this theory, 'I' is equivalent to 'this very act's doer'. But:

If we start with a separate pre-existing I, and a reflexive reference to 'this very act's doer', there will be the problem, familiar by now, of the nature of my knowledge that I = this very act's doer. I know it by doing the act, but how do I know that the pre-existing I is the doer of that reflexive act? (ibid. p. 87).

Well, 'the problem' may be familiar to Nozick, but he has not taken sufficient trouble to make it familiar to his readers. What exactly is the problem now? By this point we are in a position to identify

three contenders:

(i) (Suggested earlier - see p. 118.) How can reflexive self-reference or self-knowledge be explained non-circularly, i.e. without recourse to the 1st-person?

(ii) What is the explanation of IEM?

(iii) What is the explanation of the guaranteed reference of 'I'?

To the first, the answer may just be 'It can't' - as suggested above. It is not yet clear how self-synthesis provides the answer, in any case. That the question is the second, is suggested by Nozick's later remarks to the effect that 'a pre-existing I' is incompatible with IEM - which can therefore only be explained by self-synthesis. He muddles together the second and third question when he goes on from these remarks to say:

Only an object synthesised by the act of referring is guaranteed to be hit by that act. Only a theory of such a synthesised self can explain why, when we reflexively self-refer, we know it is ourselves to which we refer (ibid. p. 90).

We have already had occasion to observe the effects of conflating IEM and 'guaranteed reference' (pp. 10-12); no more need be said on this mistake here. It should be clear how Nozick thinks the second and third questions can be solved by self-synthesis. The first question would be 'solved' in the following way, presumably: To refer to oneself reflexively is not to refer to a pre-existing object of which I know that it is I; but rather to synthesise such an object. (But this doesn't remove the circularity; and I think Nozick has probably dropped this question.)

What precisely is self-synthesis? Nozick vacillates. He asks whether we can say that what performed the reflexive act was the entity which the act synthesised:

Can the rabbit be pulled out of the rabbit? It is some such theory as this that Fichte presents; he speaks of the self as positing itself, also of the self as positing itself as

positing itself ... Is there really no pre-existing self independent of the act A₀ of reflexive synthesis? Can the self really be a Fichtetious object? (ibid. p. 89).

It must be so if the self-synthetic account is to be a solution to the problems Nozick has set himself; yet he draws back from its evident absurdity by talking, not wholly intelligibly, of a current synthesis affecting a later synthesis as a 'non-binding precedent':

A fresh creative act is not necessary with each act of reflexive self-reference. Usually, the self habitually follows the earlier precedents ... Sometimes self-reference will occur more self-consciously, a self will step back from habit to consider its nature (ibid. p. 89).

It seems that for someone concerned to expose the illusion of pre-existing selves, Robert Nozick is rather too knowledgeable about their activities and predilections! But the provenance of his knowledge is uncertain. He concludes his account of self-synthesis with the throwaway remark, in Humean vein, that the 'delineation' of the boundaries of the act of reflexive synthesis

might give rise to the illusion that the reference is to an independently pre-existing and bounded entity (ibid. p. 90).¹

¹ Nozick refers to Hume's 'bundle of perceptions' when he floats, as an afterthought, the view of the self as a property. In virtue of having the property (P) of being reflexively self-referring, I also have the property of being I. The implications of this view for Nozick's own conception are presented with refreshing directness:

The property of being I came into existence at a certain time in virtue of the property P being instantiated in what arose from the fertilisation of Sophie Cohen Nozick's ovum by Max Nozick's sperm, a fact for which I am very grateful (ibid. p. 113).

But the view of self as property is rejected on the, for Nozick, I would think, insufficient grounds that it 'seems

This really is the most extraordinary flight of fancy, unanchored by any kind of appeal to empirical fact or responsible philosophical argument. The theoretical apparatus of 'classification' and 'entification' attendant on the 'serious' notion of self-synthesis is a wonder to behold; it is called into being (or maybe it synthesised itself) as a result of Nozick's awareness of the obvious objection to his account. This is one made explicitly by Richards in his [1984]

... hasn't Nozick ... smuggled into the first stage of self-synthesis an underlying unity, which is both essential to the process and has a 1st-person nature? (ibid. p. 156).

This is of course the pre-existing agent which performs the act of self-synthesis. Richards believes that the account can work despite his criticism, and goes on to offer an unconvincing Nozickian supplement drawing on Castaneda's Guise Theory (ibid. pp. 158-165). Nozick himself misappropriates Lichtenberg, the first Trojan, in support of the idea that just as thinking can be going on without a subject, so can acting (ibid. pp. 87 and 93). But Lichtenberg was concerned to show what, on Cartesian grounds, one is entitled to assert about thought; if one doubts the existence of others, one is not entitled to the distinction between self and others. But this does not demonstrate the intelligibility of subjectless thought (let alone action).

Mention of Lichtenberg brings us to the heart of the diagnosis of self-synthesis. For in his espousal of subjectless action, and in his rejection of a persisting substantial self, Nozick's view may seem to have deep affinities with Trojanism. This is indeed so; but its affinities with Cartesianism are deeper. To the committed Labour Party supporter Social Democrats and Liberals are 'pink Tories'; and to the committed Greek, Trojans are 'pink Cartesians'. So the assertion of the dual affinity of self-synthesis with the two anti-Greek positions should come as no surprise.

like a bit of philosophical chicanery, too much froth and too little substance' (ho, ho) (ibid. p. 114).

Let us spell out these compressed remarks. Trojanism and Cartesianism each offer substantive metaphysical responses to the existence of the linguistic peculiarities of 'I' which have been under discussion (IEM, guaranteed reference, privileged access, etc.). Self-synthesis is clearly a similar response; and it amounts, I think, to a 'solipsism of the present moment' which is the end-result of the Cartesian way. Anscombe's objection to the Cartesian res cogitans expounded on p. 68 comes to mind; viz. that its identity persists for the duration of the thought-episode only (which is no time at all). And as noted, the trouble does not stop there; for the self required by Descartes and postulated by Nozick as self-synthesised can have no criteria of individuation as well as none of identity (see p. 68). I said above that Nozick thinks he can make play with the idea of subjectless action; but if this idea is denied him, how was it that the Trojan's espousal of subjectless thought was accorded a degree of respect? Well, on the Trojan account the subject of reflexively self-referential acts is clearly identifiable; it is a person, a persisting, embodied subject. In the case of the present writer, that subject can be referred to by means e.g. of 'Hamilton'; but not by means of 'I'. Hamilton is a subject of thoughts and experiences; but he is not referred to when he says 'I am thinking of terminating this discussion'. The Trojan view is of course plagued with difficulties - and indeed the militant Trojan (see p. 50) may have special problems in accommodating the concept of a person as a subject of experiences. But these difficulties are simply multiplied when instead of saying no object is referred to by 'I', we say that a 'self-synthesised object' is referred to. Nozick's self-synthesis is indeed a limiting case of 'a cloud of philosophy condensed in a drop of grammar'; or should it be 'a cloud of conceit ...'?

3.2) 'Fortuitous' and 'Aware' Self-Reference; the Circularity of the Self-Reference Principle

(i) Anscombe's 'striking phenomenon'

It might be thought that a good way into reflexive self-reference is via consideration of the question favoured by Anscombe and insisted on by her at some length in Elizabeth Anscombe und der Bangor Kreis (see Appendix) - viz. what is it that someone realises when he comes to know that the person he has been referring to by some name or description is in fact himself? What is it that Oedipus realises when he comes to know that it is he himself who has brought trouble to Thebes, or Emma, when it dawns on her that it is she herself who has brought about the spring in Mr. Knightley's step? Such cases can be very striking, especially when, as in the Oedipus example or those in der Bangor Kreis, the subject has no idea, before the dawning of self-knowledge, who the referent may be. Two salient grammatical facts which these cases bring out (the first has already been discussed) are these:

- (i) There is no sufficient yet non-circular account of the self-reference of 'I'.
- (ii) 'I'-statements are not reducible to statements which do not use 'I'.

The first is insisted on by Anscombe, as we have seen; the second is endorsed by Nozick (op. cit.) and Perry [1979].¹ (It might be questioned how far (ii) is acceptable to Trojanism and therefore whether it is a 'grammatical fact'; certainly it conflicts with Wittgenstein's Philosophical Remarks account. But this matter will

¹ In addition to the passage quoted above (p. 118), Nozick also discusses how 'I'-statements are not deducible solely from non-'I'-statements - the converse phenomenon to non-reducibility (ibid. p. 73).

not be pursued here.) The non-reducibility of 'I'-statements is manifested in our knowledge-attributions to subjects. If I know that I am in hospital, it doesn't follow that I know that Andrew Hamilton is in hospital - the reason for my being there might be Korsakov's Syndrome caused by a stress-related alcoholism of the protracted final stages of thesis-completion. It might even be possible that there are no identifying descriptions of myself that I know, and which therefore could be substituted for 'I'.¹

In fact, the striking phenomenon which Anscombe is excited about (der Bangor Kreis p. 238) is not peculiar to self-consciousness. Enver Hoxha, long-time and wily post-war ruler of Albania, made the horrifying discovery that his equally long-time deputy of 40 years, Mehmet Shehu, had all along been in the pay of, at various times, the Gestapo, the CIA, the KGB, MI5, and the Italian and Yugoslav secret services.² How extraordinary, then, was the discovery expressed by 'It is my trusted aide, Mehmet Shehu, who has brought trouble to Albania!' A husband becomes convinced that his wife is having an affair. Imagine his awful realisation 'It is my best friend with whom my wife is having an affair'. A shadowy figure, known only by the presumed effects of his presumed actions, suddenly stands revealed in broad daylight - as someone very close to one. It is true that he is described as 'my aide', 'my friend', so a certain range of feelings and attitudes becomes appropriate. The realisation 'It is the Chancellor of the Exchequer with whom my wife is having an affair', though also striking in its way, would not normally invoke the same feelings of betrayal. But these

¹ Nozick considers, as a possible synonym, 'this very producer of this very token' - and rejects it because there is 'no guarantee that I (reflexively) know that I produced it' (ibid. p. 80). See also p. 118 above.

² Or so Hoxha claimed. Shehu's apparently stalwart support as deputy leader in the People's Republic would be explained as 'deep cover'. Whether the discovery was in fact prompted solely by the great Albanian's paranoia is beyond the scope of the present work (see The Artful Albanian: Memoirs of Enver Hoxha, London 1986 (pp. 327-37) - a remarkable exercise in self-justification).

feelings, though self-regarding, are not prompted by a discovery about oneself, but about another; which however exhibits just that strikingness of the Oedipus example. Oedipus now refers to himself self-consciously rather than fortuitously; but do we need to say 'what his realisation consists in', as Anscombe implies, any more than we need to say what Hoxha's realisation consists in? 'They now know the identity of the perpetrator of the troubles' seems to capture the realisation in both cases.

What is striking about the self-conscious cases seems to extend to 3rd-personal discoveries too, then. Further, many self-conscious cases are not that striking - when they betoken no dramatic self-discovery. Emma discovers that she is the object of Mr. Knightley's love - as much a discovery about Mr. Knightley as about herself. But one may say here that she is finding out about love as much as that she is finding out about herself. (Those who are most struck by 'voyages of self-discovery' and the like, particularly as pertaining to themselves, generally possess the idlest and weakest minds which can find little of greater importance to dwell on.) To take a more mundane example: The present writer searched the basement of the Philosophy Department at St. Andrews University to find the source of a curious squeaking noise. Rodents perhaps? They proved elusive until he realised that he himself was making the noise - his boots had let in water under the heel and were squeaking as they dried out. A self-conscious realisation - but the most important discovery was about the properties of certain footwear.

The preceding remarks about Oedipus and Hoxha may serve to cast doubts on our original idea: that self-consciousness will be illuminated by considering the difference between fortuitous and reflexive self-reference. 'Well', it may be said, 'undoubtedly reflexive self-reference is integral to self-consciousness (though we've yet properly to characterise the former notion); but all that the examples of self-conscious realisation serve to bring out is the difference between fortuitous and "aware" reference, which happens to be to oneself but is equally manifest in the 3rd-personal examples. Consequences for his own thought and action follow from Oedipus' realisation; but equally, consequences for his own thought and action follow from Hoxha's. The possibility of fortuitous

reference follows simply from the fact that the singular terms in question have a sense as well as a reference; we have said next to nothing about self-conscious self-reference just by saying it is non-fortuitous.'

This response presupposes (to put it as vaguely as it is possible to do) that there is something 'special' about self-consciousness. This is a feeling which it is hard to deny! But it would be wrong to think that many of the constituents of self-consciousness might not, by themselves, be quite 'ordinary', and encountered in one's consciousness of others. Thus, 'reflexive self-reference' may just be 'aware' self-reference. In cashing out the latter notion, however, it will be apparent that not a lot is being said about self-consciousness. This is because, of course, 'aware' self-reference is self-reference plus the circularity-inducing requirements already discussed (pp. 58-62). What are we to make of the alleged circularity of the 'self-reference principle'?

(ii) Is the self-reference principle circular?

The principle, obvious enough to all appearances, in fact requires lengthy and careful presentation. It is fundamental to any account of self-consciousness. Although the 'no-reference' view has been abandoned, the diagnosis of the motivation for the viewpoint it expresses may still with profit be engaged in. Anscombe, we recall, is worried about circularity. 'If the self-reference principle is circular, it tells us nothing about self-consciousness. Attempts to build something substantial on its foundations are vertiginous. We must abandon it therefore - together with the Greek view it expresses'. - This seems to be her underlying view. When one examines the self-reference principle, one cannot but have some sympathy for this view. But, as Evans says, 'nevertheless [the] principle ... must be respected'.

The precise questions to be considered now are these:

(i) Is the principle in fact circular?

- (ii) If it is, what is the significance of this fact? Is the circularity 'vicious'?

Related to these is the further question:

- (iii) To what extent is the self-reference principle meant to capture the idea of self-consciousness? Would Anscombe's 'A', as well as 'I', fall under the principle?¹

The circularity in the self-referential account, it will be recalled, consisted in the following. In the definition

- 'I' is the word each one uses (a) to refer to some person,
(b) knowing and intending that person to be himself

the knowledge and intention can be specified only by using the 1st-person. What must the subject know and intend? 'That I am the person referred to (by "I")' (see p. 61). The matter of circularity is discussed by Evans in an 'Appendix' to his chapter on self-identification in his [1982], and it is to this that we now turn.

This 'Appendix' contains some of Evans' deepest and most suggestive remarks on self-consciousness. But its opening criticisms of Anscombe are not satisfactory as they stand. Evans takes her circularity objection seriously, and seeks to demonstrate non-circularity (rather than showing that circularity doesn't matter). He writes:

It is perfectly possible to ascribe to a subject the intention to refer to himself, in the sense of the intention of bringing it about that he satisfies the one-place concept-expression ' ξ refers to ξ '. Of course intending to satisfy [this] ... concept-expression is the same as intending to satisfy the one-place concept-expression ' ξ refers to me' (since 'I satisfy λx (x refers to x)' is logically equivalent to 'I refer to me'). But it does not follow that in order to elucidate the intention of satisfying ' ξ refers to ξ ', we need a grasp of the self-

¹ See pp. 138-9.

conscious Idea-type that we have of ourselves (ibid. p. 258).

He later talks of 'the self-reference principle'

... that a fully self-conscious thinker will be aware that he satisfies the concept-expression ' ξ is thinking of ξ ' (ibid. p. 260).

Well! This is perplexing stuff. The preferred intention (to bring it about that he satisfies ' ξ refers to ξ ') is not one that the ordinary speaker could articulate; it does not keep faith with any conception of his intention that he might reasonably be held to have. (The problem here is a bit like that of a speaker's understanding of the axioms of a systematic theory of meaning.) Furthermore, if the original Anscombe objection is allowed, then it must be allowed to apply here too. For the intention in question could be ascribed to Oedipus in one of his fortuitously self-referring moods. Let us say that he intends to extract a confession from 'the slayer of Laius' (not realising it is himself): viz. 'I did it'. Then, by the lights of relatively unconstrained intention-ascription licensed by Evans' self-reference principle, he intends to make the slayer of Laius self-refer. So he intends to bring it about that he satisfies ' ξ refers to ξ ' - without realising it. Therefore it is not sufficient that an agent have this intention since he may be mistaken about the means required to fulfil it.

Whether or not one feels that the foregoing insufficiency charge can be made to stick (and it will be a patient reader with a high tolerance for the absurd in philosophy who has got so far as understanding it), Evans obviously felt that a circularity charge was more pertinent. It is to this charge that his final disingenuous disclaimer is directed. But it was not suggested that the self-referential accounts were circular because they presupposed, as Evans writes, 'a grasp of the self-conscious Idea-type that we have of ourselves'; simply that the 1st-person figured in the explanans of which it is the explanandum. To overcome the insufficiency charge (and perhaps anyway) it needs to be said that the speaker intends to bring it about that he himself satisfies ' ξ refers to ξ ' - where 'himself' is the 'I'-reflexive, requiring explanation in terms of the 1st-person. So it looks as if anyway circularity has

not been transcended.

(iii) The consequences of circularity

The correct presentation of a self-reference principle in the Evans style, though now of course circular, would be as follows, then:

A fully self-conscious thinker will be aware that he himself satisfies the concept-expression ' ξ is thinking of ξ ' (where 'himself' is the 'I'-reflexive).

This principle is essentially circular; the requirement on the subject's awareness will contain, explicitly or covertly, an 'I'-reflexive.

What is the consequence of this essential circularity? It will be worrying only if one is occupied by one of the following sorts of thought (which are probably related). First, that it must be possible to furnish a completely introductory explanation of any term in the language, and in particular of 'I' - an explanation that does not presuppose any particular previous understanding. But this view, which has its origin in a network of ideas related to that of the primacy of ostensive definition, is surely rather antique, and quite implausible when contrasted with the idea that 'light dawns gradually over the whole' of language for the novitiate. Second, the circularity will be worrying if it is thought that the nature of self-consciousness must be made explicit by examining the consequences of the self-reference principle. This is a rather coy formulation, but what I have in mind is this. In the original presentation of the self-reference principle by Evans, the subject simply has to be aware that he satisfies ' ξ is thinking of ξ ' - rather than 'he himself', which incorporates the self-conscious perspective from the outset. Evans then considers various proposals which seek either to make clear what the subject's awareness involves, or what it implies (ibid. pp. 259-60). These have validity despite the circularity of the self-reference principle, and will be considered shortly, but acknowledgement of the circularity will mean that one's attitude to the proposals will be

different. This is because one will, in a sense, regard self-consciousness as something ineffable, since it is impossible to explain the phenomenon non-circularly - at least, that is how it might be put by someone in the Evans camp. One could never, for instance, explain self-consciousness as a species of knowledge of objects - because the self-reference principle, as currently viewed, does not merely say that the self-conscious subject is aware that he (an object) satisfies ' ξ (an object) is thinking of ξ (an object)'. Rather, to reiterate, it implies that the subject's awareness that he satisfies that expression is already self-conscious - a fact which ultimately we can put no further gloss on, certainly not that involving the viewing of an object in a certain way.

The conclusion that consciousness of self and consciousness of objects is each sui generis (it is not as if 'consciousness of objects' is here any more clearly understood than that of self) anticipates the denouement of the Homeric Opposition in Chapter 4 (see pp. 214-16). Certainly, if the self-reference principle is circular then we are going to have to say something like the foregoing - something involving the fact that the nature of self-consciousness cannot be made fully explicit. Clearly, there are major meta-philosophical issues arising at this point, concerned with exactly what should be involved in giving a philosophical account of self-consciousness (or anything, for that matter) - these I'll do my best to sidestep. The danger of talking vacuously is ever-present; but it does seem clear that there is at least a difference in attitude, between proponents of circular and non-circular principles, regarding proposed supplementation of self-reference. I will simply expand on these differences of attitude, without going on to assess the proposals. Evans writes:

What is required for a subject genuinely to be a self-thinker is a difficult and obscure question ... Some would say ... that it is not implausible to hold that there is an essential connection between self-consciousness and the conception of oneself as the subject of certain psychological properties: specifically, thinking, and any properties that are necessarily possessed by a thinker. According to this view, what we need to add to the description of the present case [given by the self-reference principle], to secure full self-consciousness, is more of the

same sort of thing: a disposition on the part of the subject to ascribe thoughts to himself (ibid. p. 259).

The proposal is of interest independently of the 'non-circular' view to which it is tied. But when detached from that view, it will no longer be seen to 'secure full self-consciousness', since no addition to the self-reference principle will eliminate its circularity.

Evans is in fact concerned to contrast two conflicting accounts of what the self-reference principle involves - each of which will have to be re-interpreted as giving, if correct, salient features rather than necessary and sufficient conditions of self-consciousness. The first, the dispositional account just mentioned, says that the principle involves the subject showing his knowledge of the identity between subject and object - by being disposed to ascribe thoughts to himself. Thus the fully self-conscious nature of his thought 'I am hot' would come out in his later willingness to judge 'I was hot, and I thought that I was hot' (i.e. there was something which both was hot and thought that it was hot). (Evans is also building in the thinker's awareness that he is a persisting subject of thoughts - since otherwise the required disposition could be manifested by a present-tense judgement 'I am hot and I think I am hot'.) The second account is really an 'anti-account' - the salient feature of self-conscious thought is something much more vertiginous than the dispositional account allows - perhaps something viciously auto-referential such that fully self-conscious thought is strictly impossible. One can in general understand what it is for an object of a thought to be its subject. But this is not enough; one must also realise what it is for any particular thought of one's own to have an identity of object and subject, one must grasp that the object of this very thought¹ is its subject.

It is doubtful, in fact, from what we have said of Nozick's 'auto-referential' account, that such a vertiginous 'anti-account' is plausible. Be that as it may, I hope I have illustrated some of

¹ Evans intended a full discussion of this last, 'vertiginous' thought; but that just outlined is all we have to go on (see ibid. p. 26ln6).

the effects of the circularity of the self-reference principle on accounts of self-consciousness, and have defused objections that such circularity is vicious.

3.3) IEM and Self-Consciousness

Having ploughed through an exhaustive process of refinement of the notion of IEM, the reader may be disappointed to discover that little is to be said about the notion beyond its dialectical function in the Homeric Opposition. This is due simply to the limitations of space caused by the need to complete the second and final stage of that Opposition - for the notion of IEM looks as if it is central to any account of self-consciousness. Something will however be said on the relation between these concepts, within the context of Anscombe's discussion of self-reference.

Anscombe has a laconic and suggestive discussion of a self-referential term 'A' which is not expressive of self-consciousness.¹ Her characterisation of 'A' is incomplete (and indeed she has confided to the present writer that she is 'not happy' with it²). But it is intended to cast some light on the nature of self-consciousness. It is worth enquiring whether, even given our rejection of the 'no-reference' view, it may nonetheless do so. Here is what Anscombe says:

Imagine a society in which everyone is labelled with two names. One appears on their backs and at the top of their chests, and these names, which their bearers cannot see, are various: 'B' to 'Z' let us say. The other, 'A', is stamped on the inside of their wrists, and is the same for everyone. In making reports on people's actions everyone uses the names on their chests or backs if he can see these names or is used to seeing them. Everyone also learns to respond to utterance of the name on his own chest and back in the sort of way and circumstances in which we tend to respond to utterance of our

¹ The discussion does, for example, suggest some thoughts to Noonan in his [1979] - not all of them quite coherent, however (but see p. 138 below).

² Appendix p. 242.

names. Reports on one's own actions, which one gives straight off from observation, are made using the name on the wrist. Such reports are made, not on the basis of observation alone, but also on that of inference and testimony or other information. 'B', for example, derives conclusions expressed by sentences with 'A' as subject, from other people's statements using 'B' as subject.

Thus for each person there is one person of whom he has characteristically limited and also characteristically privileged views: except in mirrors he never sees the whole person, and can only get rather special views of what he does see. Some of these are specially good, others specially bad. Of course, a man B may sometimes make a mistake through seeing the name 'A' on the wrist of another, and not realising it is the wrist of a man whose other name is after all not inaccessible to B in the special way in which his own name ('B') is. (ibid. p. 49).

She goes on to say

the reidentification of selves ... is not any part of the role of 'I'. [But] the corresponding reidentification was involved in the use of 'A' ... (ibid. pp. 52-3).

'A' exhibits guaranteed reference (1) (see above, p. 13), but not (2), Anscombe implies:

The 'A'-user means to speak of a certain human being, one who falls under his observation in a rather special way. That person is himself, and so, given that he has grasped the use of 'A', he cannot but be speaking of a real person ... [But] we saw that the 'A'-user would not be immune to mistaken identification of someone else as 'A' ... The suggestion [that 'I' would not] seems absurd ... (ibid. p. 57).

In fact Anscombe is confusing guaranteed reference with IEM at this point. Both 'A' and 'I' exhibit guaranteed reference (diagnosed as uninteresting on pp. 13-15). Some uses of 'I' exhibit IEM; which 'A' never exhibits. This indeed seems to be the crucial difference between them; and since, as Anscombe maintains, 'I' manifests self-

consciousness but 'A' seems not to, IEM would seem to have something important to do with self-consciousness.

Anscombe's discussion is far from clear, however, as well as being incomplete. 'A'-users lack the capacity to make IEM-judgements, a fortiori avowals. They therefore cannot articulate intentions, and thus cannot be said to act, merely to 'behave'. It is not, then, strictly their 'actions' on which they 'report', in bizarre fashion, using 'A'. By means of 'A', it would seem, they can make non-IEM judgements (Wittgenstein's 'I'-as-object use - see p. 3 above), plus perhaps some non-IEM caricature supplanting our IEM-judgements (the 'I'-as-subject use). Noonan, in his [1979], argues that the capacity to use 'I'-as-object requires a capacity to use 'I'-as-subject; but he sees the former as distinct from 'A'. However, Anscombe in der Bangor Kreis, in reporting her dissatisfaction with her earlier discussion, suggests that the use of 'A' may depend on understanding of, inter alia, 'This is my body' (below, p. 242). A doubt is thus raised over whether, even as it stands, the story of the 'A'-users does not covertly depend on their having an understanding of 'I' and in particular the 'I'-as-subject use.

It is not in fact Anscombe's intention in this story to expand on the connection between IEM and self-consciousness - especially given she is not clear about IEM! As already discussed in Chapter 2, she seems rather to be directing her discussion at the 'misleading platitude' that self-consciousness is consciousness of the object one is. For the reason there given, this attempt is wayward (p. 63 above). It is, however, worth exploring one line of argument intended to yield the down-grading of the 'self-reference principle' - thought not quite in the Trojan way which Anscombe intends. This is (briefly) as follows:

(1) 'A' is not expressive of self-consciousness.

(2) 'A' falls under the self-reference principle.

so (3) The self-reference principle does not illuminate self-consciousness.

Now, the self-reference principle as thus far formulated employs the 'I'-reflexive 'himself'; so clearly 'A' will not fall under that principle. However, there is a self-reference principle under which it falls, sufficiently analogous to give support to the conclusion of the argument above. It is the following:

'A' is the word each one (a) uses to refer to some person,
 (b) knowing and intending that person to be himself
 ('A*'-reflexive).

The 'A*'-reflexive (so-called to distinguish it from our earlier 'A'-reflexive, pp. 60-62 above) is meant to convey the fact that the 'A'-user must know: 'A is the person referred to'.

But really, this argument is so hedged about with qualifications that it is doubtful that much can be made of it. It is not clear what is conveyed by the 'A'-version of the self-reference principle; what does the 'A'-user understand by 'A is the person referred to'? Even if this were clear, and the principle correct, is mere analogousness to the self-reference principle for 'I' sufficient? And what does it mean to say self-consciousness is not therefore 'illuminated'? We have encountered this line of thought (that self-consciousness must be 'special', cannot be contaminated by principles operating elsewhere) before (p. 129).

Likewise it is not clear how far the Anscombe thought-experiment elucidates the connection between IEM and self-consciousness. This connection is clearly a topic for a future work.

CHAPTER

FOUR

In this final chapter the second stage of the Homeric Opposition is discussed and, as far as possible, resolved. Although the problem of avowals which is its subject-matter is in many respects of great difficulty, it is in fact easier to be confident that some sort of solution has been achieved than in the case of the first stage of the Opposition. This is simply because the issue is ultimately (if not perhaps at first) much clearer! The more closely we looked at 'no-reference' Trojanism, the harder it became to discern just what the position amounted to. In contrast, the Trojanism of the non-assertoric thesis of avowals addresses a real issue that compels some response, and is as a result easier to define - even if in the end it is rejected.

The opening section provides a brief resume of the so-called 'non-assertoric thesis' of avowals as it appears in the work of Wittgenstein and Ryle, and of the problem which provoked it. Next, detailed work on selected examples of types of avowal is undertaken, in an attempt to display the diversity of the phenomenon. This work is necessarily incomplete, but at least the concentration on the 'pain-paradigm' is mitigated. An account of an argument to the non-assertoric thesis follows, and the final and longest section discusses the Wittgensteinian idea that avowals are primarily expressive, and whether this apparently (but only apparently) Trojan proposal offers a resolution of the problem. It is contended that the notion of 'expression' is indeed at the centre of such a resolution; but that it need not imply that avowals are not assertions. The 'non-assertoric thesis' of which the 'expressive thesis' is often seen as a variant, in fact faces serious difficulties from the existence of truth-value links, which are discussed. But furthermore, provided we give the right account of the cognitive and hence authoritative status of avowals (as not involving some analogue of 'expert' knowledge), we do not need a non-assertoric account. Hence the account proffered here involves a progressive refinement of the Trojan view - a refinement out of existence. An analogous account of authority replaces the non-assertoric and non-cognitive theses, and that authority is explained

conceptually by the idea of a subject of experience giving expression to its avowable states.

Such an account is broadly Wittgensteinian, except that Wittgenstein himself clearly advocated a non-cognitive account of avowals. The extent to which his remarks imply a non-assertoric (as opposed to an 'expressive') thesis is in fact unclear, but at the outset of the discussion the traditional view (that they do) is taken, and the subtleties in the various viewpoints are gradually introduced.

4.1) The Problem of Avowals

The two most well-known accounts of avowals are those of Ryle and Wittgenstein. The latter hardly offers an account in fact, more a series of suggestions that avowals should be viewed in a certain way - as akin to purely expressive behaviour. His most categorical statement occurs in Zettel :

Plan for the treatment of psychological concepts. Psychological verbs characterised by the fact that the 3rd-person of the present is to be verified by observation, the 1st-person not. Sentences in the 3rd-person of the present: information. In the 1st-person present: expression ((Not quite right)).
(Z 472)

(It will be considered shortly why this is 'not quite right'). In the Investigations, we are several times encouraged to view the avowal that one is in pain as akin to natural, unlearnt, purely expressive pain-behaviour. But as in the Blue Book discussion of IEM, various contexts exhibit an ambiguity in Wittgenstein's account: is the claim (i) that 'I am in pain' is not an assertion about a particular person, or (ii) that it is not an assertion at all? (an ambiguity noted by Hacker in his [1972] p. 263).

Wittgenstein writes:

... in saying ['I am in pain'] I don't name any person. Just as I don't name anyone when I groan with pain. (PI 404)

and

But surely what you want to do with the words 'I am ...' is to distinguish between yourself and other people'. Can this be said in every case? Even when I merely groan? (ibid. 406)

If these passages are intended to support a merely 'no-reference' view of 'I' the comparison with groans of pain is unsuitable, as already noted in connection with the passage on IEM in the Blue Book

(p. 67). (Admittedly, other examples of sentences with a pseudo-subject which may serve as analogies are hard to find. Wittgenstein has already used 'It is snowing' in the Philosophical Remarks.) The suggestion is less oblique in the following passage from Part II:

A cry is not a description. But there are transitions. And the words 'I am afraid' may approximate, more or less, to being a cry. They may come quite close to this and also be far removed from it ... But if 'I am afraid' is not always a cry of complaint and yet sometimes is, then why should it always be a description of a state of mind? (ibid. p. 189).

The existence of this continuum opens up the possibility that a non-assertoric account of the 'expressive' uses of e.g. 'I'm afraid' will leave unresolved the problem of what to do with the 'descriptive' uses. Perhaps they should be treated as 'non-avowals' in the sense of the Famous Five - whatever account one chooses to give of the central cases of avowal. Their existence may be the reason why Wittgenstein felt the Zettel account above to be 'not quite right'. (Or perhaps he felt the terms 'information' and 'expression' were 'not quite right' - the account will be taken up at length below.)

Ryle too is opposed to the idea that avowals are primarily self-descriptive, and also favours a 'non-assertoric' account. He too is not unaware of the difficulties with such an account, but is prepared to gloss over them:

If the avowal ['I feel depressed'] is to do its job, it must be said in a depressed tone of voice; it must be blurted out to a sympathiser, not reported to an investigator ... if we are suspicious [of the avowal] we do not ask 'Fact or fiction?', 'True or false?', 'Reliable or unreliable?', but 'Sincere or shammed?' ([1949] p. 99)

And later

[The] grammar [of avowals] makes it tempting to misconstrue all the sentences in which they occur as self-descriptions. But in

its primary employment 'I want ...' is not used to convey information, but to make a request or demand. (ibid. p. 175)

Talk of an avowal 'doing its job', or of the 'primary employment' of an expression, is an attempt to focus attention away from those awkward instances where an avowal does indeed seem to have a descriptive role.

A consequence of the non-assertoric thesis is the non-cognitive thesis of avowals; one stated explicitly by Wittgenstein, but grounded by him in the principle that knowledge requires the logical possibility of doubt, rather than on the non-assertoric thesis. It is the possibility of a non-cognitive thesis of avowals that prevents us from supplementing feature (v) of the Famous Five (on the impossibility of doubt) by saying that knowledge is guaranteed. Hacker, in presenting the non-cognitive thesis, refers to Wittgenstein's advocacy of the impossibility of self-knowledge. This paradoxical phrase is acceptable provided one realises that it does not imply any epistemic inadequacy - that is comparable to 'the impossibility of feeling the pain of another'. (The non-cognitive thesis is discussed below (pp. 208-11).)

Now, what is it that prevents us from regarding the non-assertoric and non-cognitive theses of avowals as interesting but idiosyncratic and ultimately unsatisfactory adjuncts to the Wittgensteinian and Rylean philosophies? Gareth Evans writes:

[Wittgenstein] encouraged us to ... think of [1st-person psychological statements] as unstructured responses to situations. (He was well aware that this would enable him not to think about certain issues.) ([1982] p. 209n)

Behind this cryptic comment of Evans seems to lie the suggestion that the non-assertoric thesis enables Wittgenstein not to think about the basis for the authoritative nature of avowals. It is this issue that is central to the second stage of the Homeric Opposition. Why is it that the subject is accorded authority-status on his avowable-states? What is the explanation for the Famous Five? We might express the problem as follows: If avowals are construed assertorically, there will be some condition which makes the avowal

true or which warrants the assertion. A dilemma then arises: either the condition is the obtaining of some inner state of the subject, to which he has access in a way denied to others, or it is some outer state (behavioural or physical) to which he has access in the way others do. In any event, his authority is regarded as being based on access to some state of himself.

Before discussing this dilemma, I need briefly to raise a question about formulation. Is it strictly the subject's 'authority' that requires explanation? Such a term seems to imply an assertoric interpretation of features (iii) and (iv). 'Avowals are authoritative' is perhaps ambiguous between 'the utterance of an avowal, if comprehending and not intended to deceive, settles the matter of the utterer's mental state', which is acceptable, and 'Avowals express the subject's authority on what his own mental state is', which is controversial. Controversial, because authority on a certain matter seems to imply a knowledge of that matter, something the non-assertoric account of avowals is intended to exclude. There is a second reason for not framing the problem in terms of authority. For isn't it the fact that the subject has different access to his mental states from that of others that requires explanation, not that his access is superior? Or what amounts to the same thing given the implausibility of 1st-person behavioural access, that he has any access at all? Re-expressing this point in our neutral fashion, the question becomes, 'Why are avowals evidence of the subject's mental state at all?' (Of course, such a question is not intended to express incredulity!) In practice, we find it impossible to conceive of the subject's access (if any) as different but not superior, since the image of 'access' is a Cartesian one, but it is worth separating the issues. It is with this practical point in mind that I continue to formulate the problem of avowals in terms of authority.

To return to the dilemma, the latter option, that the truth-or assertion-condition of the avowal is the obtaining of some outer behavioural state, is quite implausible, though maintained e.g. by Carnap [1959]. It is in contradiction with the first feature of the Famous Five, and leaves the explanation of the rest quite obscure.

In contrast, the former option preserves all the Famous Five

features whilst encouraging an unacceptable Cartesian explanation of them. The authority which an individual has in pronouncing on his own mental state by means of the appropriate avowal (if this is how it should be described), and which is expressed in features (iii) and (iv), is explained in terms of the subject's superior access to an inner state. The especially direct character of this access explains features (i), (ii) and (v). But for two reasons which are well known, this is the sort of explanation we can do without. Taking these in the correct historical order: First, it was realised that this Cartesian position led ineluctably to sceptical problems over other-ascription - the so-called 'Problem of Other Minds'. (It is perhaps worth noting that the sceptical problems thus generated may not apply in the case of those psychological predicates whose self-ascription does not involve any special authority of the subject, e.g. 'is intelligent', 'is excitable'.) Second, it was shown by Wittgenstein in the Investigations that Cartesianism is not secure even within some 1st-personal redoubt, because of the apparent incoherence of the idea of a logically private language which it seems the Cartesian is forced to embrace. He faces a 'Problem of Non-Other Minds' too, then (to use R.C. Buck's phrase from his eponymous [1962]).

It is the apparent dilemma between inner and outer truth- or assertion-conditions that the non-assertoric thesis intends to undercut. We shall assess later the validity of this dialectic; but one further apparent option should be disposed of from the outset. This is to claim that the authority of avowals is a 'mere convention' in no need of explanation; and to leave things at that. But to try to say this is to say that any other condition would do just as well (e.g. we could cancel my authority for my own mental states and say that my best friend has authority for them). And this is something we do not feel. An explanation for the authority functions as a kind of justification for it.

4.2) Avowals of Belief(i) The categories of avowal

The treatment of avowals thus far has been highly general and abstract; nothing has been done by way of applying the Famous Five features to particular 1st-person present-tense psychological utterances and thus delimiting the class of avowals. Which psychological self-ascriptions in the present tense actually fall into the category 'avowal'? (The concentration in the literature on the case of 'I am in pain', coupled with apologetic excursions into the realms of 'I have a pain in my foot' etc., is quite remarkable.)

What counts as psychological, rather than physical, self-ascription, will depend upon one's purposes. (Though the traditional distinction between mental and physical is, as Gareth Evans has remarked, 'almost entirely arbitrary' ([1982] p. 223).) What is clear for present purposes, however, is that many traditionally psychological self-ascriptions do not satisfy the Famous Five and are therefore not avowals. These are self-ascriptions of more enduring qualities (dispositions) that can be made on a behavioural basis and for which the subject is, to use a tendentious phrase, in no better position than others well-acquainted with him to make the ascription. They are thus akin to 3rd-personal ascriptions. Such will include those of character traits, moral qualities and pathological states. These carry a heavy predictive content which furnishes the possibility of falsification in a way not catered for by the Famous Five.¹ At the other end of the spectrum will be the 'paradigmatic' avowals of episodic mental states - sensations, perceptual appearances, moods, feelings and emotions. In their lack of propositional content these states will contrast with a third category of self-ascriptions which occupies an intermediate position on the avowal / non-avowal continuum - states which have propositional content and which figure

¹ Though see pp. 227-29 below.

in the self-ascription of propositional attitudes. These I term 'qualified' avowals.

The sub-category 'paradigm' avowal is, as the term suggests, the sort of avowal that most readily springs to mind when that kind of self-ascription is under discussion. These are avowals of mental episodes which have natural behavioural expression - thus affording a range of linguistic and non-linguistic behaviour which furnishes criteria for the mental state. It seems right to say with Rorty [1970] that the ascription of these mental states is compatible with any range of future behaviour of the subject (see pp. 162-3). Thus it is plausible to hold that the holistic constraints, which complicate the ascription of those states which feature in 'qualified' avowals, are absent here.¹

Examples of 'paradigm' avowals would be 'I am in pain', 'I am angry', 'I am nervous', 'I am joyful', 'I am afraid', 'I'm hot', 'I'm cold', 'I feel sick', 'I feel depressed' ... i.e. we are dealing here with sensations and with a range of feelings, emotions and moods.

(ii) Sources of error in avowals of belief: identification and attitude

What this range is, is discussed below (p. 178). Our concern now is with present-tense self-ascriptions of propositional attitude. To what extent do they conform to the Famous Five and thus belong to the class of avowals? Let us break down the ascription into the familiar identification and predication components and locate three contending sources for error. (I will deal generally

¹ These constraints do not serve to over-ride individual self-ascription in particular cases - except perhaps where the 'qualified' avowal is corrigible (see pp. 153-56). Rather, if an individual's self-ascriptions fail to conform to the constraints in terms of presenting a satisfying picture to others, that individual would be held not to be competent in self-ascription.

with all sorts of propositional attitude at this point, though focusing on belief. Desire I will deal with later.)

First the identification component. No error is possible here; it makes no sense to ask 'Someone believes that p , but is it I?' In general, it seems that all propositional attitude self-ascriptions in the present tense are IEM.

Next the predication component. This can be divided into the attitude element, and the propositional element. With regard to the former, it might seem that the question 'I have a certain attitude towards p , but is that attitude one of belief, or doubt, or some other attitude?', is intelligible. But I cannot, for instance, wonder whether my attitude towards p is one of belief or, say, desire (though I may wonder whether I believe some important fact simply because I very much want it to be so). To borrow a phrase from Mark Platts, one attitude can be substituted for another only if it at least has the same 'direction of fit' with the world (see his [1979] pp. 256-7). More importantly, the formulation is tendentious. How? Well, as Evans has pointed out, in making a self-ascription of belief (to limit ourselves for the moment to that attitude), I consider the question whether p , not whether I believe that p ([1982] p. 225 - see below). It may be thought, then, that questions about what my attitude is can be translated into questions about the propositional element (see below); that is, if I am wondering whether my attitude towards p is one of belief or doubt, I am in fact wondering whether it is p or $\sim p$ that I believe. But as may be guessed, questions about which shade of belief I have (from commitment to increasingly tentative evaluation) cannot be so treated. In fact, Evans' point has direct application only in the discussion of the propositional element; it is an analogous point that needs to be made here. That is, that the formulation is tendentious because it misrepresents the business of arriving at an attitude as a process of investigation into something that was there all along.¹ This is brought out by the consideration that once the sincere comprehending avowal 'I believe that p ' has been made, it makes no sense to wonder whether it was belief that was my attitude.

¹ cp. Hacker [1986] p. 302.

In contrast to the case of the identification component, then, there is a temporal restriction on the unintelligibility of the doubt.

With attitudes other than belief, however, the doubt can on occasion make sense, because of two caveats. Firstly, there is the question of the long arm of the Unconscious, to be discussed more fully shortly. A person may avow¹ that he fears that his father will suffer a violent death, but under analysis can be brought to see that this apparent avowal was mistaken; for he desired it. However, analogously to the point made about the reflection prior to avowal, this post hoc error can justify the substitution of a propositional attitude from a limited set only; the original attitude (fear) can be replaced only by its polar opposite (desire). Secondly, a caveat analogous to the de re caveat pertaining to the propositional element (see below). On that matter Andrew Woodfield writes:

... a de re thought ... has an external aspect which consists in its being related to a specific object. Because the external relation is not determined subjectively, the subject is not authoritative about that. ([1982] p. viii)

Analogously, a subject may avow 'I know that p', but be mistaken - 'p' is false and he merely believed that p. Philip Pettit has argued that this dependence of the propositional attitude on the 'non-subjective' i.e. on aspects of the world as well as of the subject, extends beyond knowledge, its traditional resting-place, and infects other propositional attitudes including belief.² Be that as it may, other cases than knowledge can be recognised on grounds other than the Wittgensteinian ones Pettit is arguing for. Take the distinction between remorse and regret. I may express remorse over the suffering someone is undergoing, because I feel it

¹ The reader can see where by 'avow' I mean 'sincerely and comprehendingly avow'.

² He calls such states 'impure', and appeals to Wittgenstein's rule-following considerations. Amplification awaits the result of the search for the text of his paper, delivered at St. Andrews in 1983. (Though now see his [1986].)

was partly my responsibility. Others persuade me that I was not in any way responsible; I do not then feel remorse, but merely regret (which implies no responsibility).¹ Perhaps I had all along felt merely regret. It is not clear that we would say this however; one might talk of irrational feelings of remorse (just as one speaks of the irrational feelings of guilt commonly experienced by the bereaved).

In the case of the propositional attitudes other than belief, then, we may be inclined to admit the intelligibility of the doubt about attitude; but, bearing the Evans point in mind, we will not be led to exaggerate the extent of the kind of investigation it licences.

(iii) Sources of error: Propositional element

Can I ask the question 'I believe something, but is it that p or that q, or some other proposition?' (We'll stick with belief just now, before broadening the discussion to treat other attitudes which may not be analogous in their behaviour.) Setting aside the

¹Theodore Strelski, who spent 19 years unsuccessfully pursuing a doctorate in mathematics at Stanford University and concluded by murdering his professor in protest, made a related point in his famous statement on release from jail:

If I committed a murder to criticise Stanford, if I express remorse, I just throw my whole argument in the waste-basket ... I feel regret. Not remorse ... regret as I see the tragic impact on people ... But if I had it to do over again, I'd do it the same way. (Trenton Times Nov. 1985)

Strelski's logic, if not his reason, is impeccable. What he regrets are the causal consequences, in terms of individual suffering, of what he sees as a justified action. He is responsible for the murder, but as it was not wrong, feels no remorse; its causal consequences (suffering) are mere natural concomitants, for which he is not responsible - they thus provoke only regret.

tendentious interpretation of this question, which implies investigation into a pre-existing belief, there are in fact two sorts of case in which the question makes sense (and thus two sets of case when honestly mistaken 'avowal' is possible):

- (i) When de re beliefs are involved - analogously to the discussion under the heading of the 'attitude element'.
- (ii) When the Unconscious is implicated - likewise analogously.

Other apparent cases turn out not to be such and it is well to distinguish these at the outset. Let us suppose I am a loyal Tory party supporter, and I say (sincerely) 'I believe that Mrs. Thatcher did not mislead the House over the Westland affair'. Someone later convinces me that this statement was a mere expression of partisanship and wish-fulfilment. Do I then say 'So I didn't believe it after all'? Rather, I think, one would say that the grounds for my belief had been inadequate, that the belief was caused more by a wish than by the cool appraisal of evidence, etc.¹ Another example: Say I profess the belief 'Socialism is the best system'. Later, in the face of repeated attacks on my socialist credentials by 'fellow-socialists' (my support for 'Solidarity' and NATO, distrust of Fidel Castro, etc. - the usual deviations), I begin to wonder whether I really am, and really have been, a socialist. Was I then mistaken about my beliefs - mistaken that I believed that socialism is the best system? There are a number of options here - but none of them adds up to the required straightforward 'honest mistake'. (Indeed it is hard to see what this requirement is - which is disturbing, since if the search is for the unobtainable perhaps the examples have been set up wrongly. But I demur on this!) I was perhaps mistaken about the implications of my socialist belief - that it conflicted with other, more important values to which I am attached, like liberty. Or maybe (and perhaps this comes to the same thing) I was mistaken about what 'socialism' meant. (Though a schoolgirl who thought social workers were socialists by virtue of semantics rather than

¹ Those who still think it plausible that I was mistaken that I believed that p will want to say that it was just the avowal that was so caused.

inclination - she had muddled the two terms - might have professed a belief defective through misunderstanding of meaning and not implications). Finally, I might say that I had been deceiving myself - I very much wanted to believe in socialism (because otherwise, I feared, I would be betraying my family background) and this desire caused my belief, but really my reasons were insufficient. (This option is analogous to that of the Thatcher example, and will be pertinent in the discussion of the Unconscious shortly.)

So care is needed in isolating the genuine examples. The first category are de re beliefs, where the subject's authority on what he believes is curtailed (see the Woodfield quotation on p. 151). For me to have the belief may require the existence of an object which does not in fact exist ('I believe Smith's murderer is insane' - see Wright [1983] p. 99). Or I may express a belief which is pragmatically self-defeating ('I believe that I'm disembodied', on some accounts). In each case, my belief is 'lost' - though, it may be said, this content-less mental state is adequately 'expressed' (no more and no less) by the words I used.

The second category involves the Unconscious. The intrusion of the Unconscious will affect any member of the range of propositional attitude self-ascriptions, so it is well to make some general comments here. Authorities disagree on how to characterise the consequences of Freudian theory for the Cartesian picture which causes our troubles. MacIntyre in his enjoyable [1958], notes the obvious restriction on self-knowledge, but goes on to claim that:

The Cartesian philosophical tradition, mediated by Brentano, reinforces and is reinforced by [Freud's] depiction of the unconscious in terms which he had elaborated to deal with the entities of neurophysiology [viz. in substantial but non-physical terms] ([1958] p. 46)

In contrast, Flanagan in his [1984] sees Freud's discoveries, if such they are, as quite destructive of the Cartesian view:

If ... introspection is highly inferential [i.e. self-knowledge is not really 'introspective' as this term is generally understood] ... or, alternatively, if introspection is

non-inferential but highly unreliable (imagine that the mind's eye has the same degree of accuracy as the real eyes of a person who has defective vision), then 1st-person psychological reports have no special status ... Freud's philosophy of mind ... implies that all our ordinary 1st- and 3rd-person methods for gathering psychological data are about equally likely to yield truth or falsity ... ([1984] pp. 67 and 73).

This latter is clearly too extreme a view; but Flanagan and MacIntyre need not be in direct conflict, in fact. To the extent to which the substantiality of the mind (i.e. dualism) is central to Cartesian doctrine, Freud is reinforcing it; to the extent that Descartes requires the activities of that substantial mind to be transparent to the subject, Freud is undermining it. However, unless the quite implausible thesis of 'The All-Pervasiveness of Neurosis' is correct, there will still be a wide range of avowals the authority of which is left unimpaired since they are unaffected by unconscious motivation. And our problem of explaining this authority will remain.

The kind of example I have in mind in the case of belief is as follows: A woman develops a quite disabling agoraphobia, and in the course of analysis the reasons for her behaviour are probed. The analyst surmises that the condition has developed as a response to her husband's jealousy and possessiveness, manifested in aggressive interrogation after every excursion she made from the house (a not uncommon kind of cause of the phobia apparently - which undermines the value of behaviour therapy in such cases). He asks her whether she believes that if she leaves the house, she will inevitably be subjected to an unpleasant interrogation on her return, and whether this is a reason for her behaviour. The patient (sincerely) disavows this belief, and reiterates her original inadequate explanation: 'I believe it's because of all those gossiping neighbours; I never know when I'm going to meet one of them in the street - it's got nothing to do with my husband!' However, as analysis proceeds she comes to see the correctness of the proffered explanation; and to believe that her earlier avowal had been mistaken. The facts, then, are these: There was a sincere comprehending avowal 'I believe that p'. At the same time, certain behaviour was rationalised by, inter alia, the belief that ~p, a

belief that was subsequently avowed. Was the subject therefore mistaken in the earlier avowal 'I believe that p'? Did she then, in fact, believe that \sim p? I would be sympathetic to those who say 'yes', whilst admitting that possibly the complexities of this kind of case (very close to those of self-deception) could lead one to say: 'There's no definite answer. The ascription of belief requires the confluence of avowal and behaviour, and does not cater for cases like this where the two seem to come apart'.

The latter view prompts the following thought: 'We do not need to appeal to "the look within" to explain the authority of avowals. For what happens is this. A subject makes an avowal. This of itself is not authoritative until or unless we believe it to be sincere; and we test for sincerity by examining the rest of the subject's behaviour. "I am in pain", for instance, guarantees the truth of "He is in pain" said of the subject only when the rest of the subject's behaviour tallies with the avowal and the latter can therefore be regarded as sincere. What causes the perplexity in the case of unconscious belief is that the behaviour expressly does not confirm the sincerity of the original avowal, and yet the latter is, despite this, claimed to be sincere'. Would it were that simple! Unfortunately, we do not, in general, 'test for sincerity'. Avowals are 'positive presumptive' - they are assumed to be sincere unless shown otherwise. Their authority is not conditional, but a prima facie authority defeasible in certain ways (see pp. 224-25 below). The preceding way of explaining their authority is thus mistaken.

(iv) Evans' account of 'belief self-ascription'

To return now to the main track of our discussion. It appears that, for a range of propositional attitude self-ascriptions of which belief is the paradigm, conformity to the Famous Five and consequent avowal status is marred by the special case of the Unconscious alone and complicated by the existence of beliefs de re. (The former exceptions might be treated as non-avowals in the sense of our second category of mental states given at the outset (p. 148).) Now, we have already had occasion to refer to Evans' remark on belief (p. 150). Does his account as a whole serve to render the problem of avowals insofar as it applies to belief

self-ascriptions more tractable? Evans' aim, in this section as elsewhere, is, to put it crudely, to undermine the Cartesian picture of self-knowledge without endorsing the Wittgensteinian one, whilst keeping the whole bathed in a Kantian light. Less poetically he boldly predicts:

I shall quite avoid the idea of this kind of self-knowledge (of what we believe and what we experience) as a form of perception - mysterious in being incapable of delivering inaccurate results (ibid. p. 225)

and goes on to quote the extraordinary remark, so compressed and resonant, reported of Wittgenstein in discussion:

If a man says to me, looking at the sky, 'I think it is going to rain, therefore I exist', I do not understand him. (Evans' emphasis.)

Insofar as one can understand Wittgenstein's remark (!), it is intended to undermine the idea that knowledge of one's mental properties involves an inward glance, Evans claims - and with this project, though not with any non-assertoric denouement, he is in accord:

The crucial point [in Wittgenstein's remark] is the one I have [underlined]: in making a self-ascription of belief, one's eyes are, so to speak, or occasionally literally, directed outward - upon the world.

It is worth quoting Evans' further remarks in full, since they have an important bearing on the question of the role of doubt in the self-ascription of belief - and hence the conformity or otherwise of such self-ascriptions to the Famous Five. He begins with an important principle:

I get myself into a position to answer the question whether I believe that p by putting into operation whatever procedure I

have for answering the question whether p.¹ (There is no question of my applying the procedure for determining beliefs to something, and hence no question of my possibly applying the procedure to the wrong thing.) If a judging subject applies this procedure, then necessarily he will gain knowledge of one of his own mental states: even the most determined sceptic cannot find here a gap in which to insert his knife (ibid. p. 225).

This is a forceful statement of a viewpoint that commands attention. Before elaborating on it, it is worth spelling out the Kantian aspect of Evans' account mentioned above, which is contained in his treatment of what more is needed in an account of self-ascription of belief over and above the principle stated. In order to make such self-ascriptions, he says, my use must conform to the Generality Constraint - principally, in this case, I must be able to other-ascribe. Otherwise

we secure no genuine 'I think' ('think that p') to accompany [one's] thought ('p'): the 'I think' which accompanies all [one's] thoughts is purely formal (ibid. p. 226).

In making the related point in connection with the self-ascription of perceptual experience, Evans writes:

I believe we have here an interpretation of Kant's remark about the transcendental 'I think' which accompanies all our perceptions. Without the background [of Generality] we have at most a formal [i.e. empty] 'I think' (ibid. p. 228).

(Kemp Smith translates as 'representation' what Evans has as

¹ Some have thought that this sentence implies that having 'got myself into a position' to decide whether I believe that p by 'putting into operation' whatever procedure I have for deciding whether p, I still have some more work to do from that position. Evans' following sentence makes it clear that, any infelicity of expression notwithstanding, this is not what he means to say!

'perception' (Critique of Pure Reason B 131-2); Kant's point, whatever it is, is not one limited to the self-ascription of perceptual experience.) In neither case, Evans claims, does the addition of the background encourage the myth of the 'inward glance'. Rather, it replaces it, it seems; in stating what more I need to be able to do in the self-ascription of belief beyond implementing the procedure to decide whether p, one should appeal, inter alia, to capacities to other-ascribe rather than to any unacceptable myths about self-ascription.

(v) Wittgenstein's account: 'Moore's paradox'

In its main features Evans' account bears a close similarity to that offered by Wittgenstein in Part II of the Investigations (pp. 190-192). It is worth looking at this account. In his discussion Wittgenstein makes frequent reference to Moore's paradox. This is not the paradox of how such a naive and unglamorous individual could become a famous philosopher, as is sometimes thought, but rather consists in the following.¹ The expression of ' \sim p but I believe that p' is taken to be pragmatically self-defeating, or self-contradictory, whereas someone may assert ' \sim p but Hamilton believes that p' with perfect propriety. Why is this?

Wittgenstein expresses the paradox in several different ways, all along trying to undermine the picture of belief which finds the asymmetry in question puzzling. He writes:

Moore's paradox can be put like this: the expression 'I believe that this is the case' is used like the assertion 'This is the case'; and yet the hypothesis that I believe this is the case is not used like the hypothesis that this is the case.

and

¹ Even the gripping narrative power revealed in Moore's racy autobiography in Schilpp ed. [1952] unfortunately does little to undermine the harsh assessment implied here.

If there were a verb meaning to 'to believe falsely', it would not have any significant 1st-person present indicative (ibid. p. 190).

Wittgenstein's diagnosis of the puzzlement this asymmetry may occasion parallels Evans' treatment. The idea both philosophers are concerned to combat is that I may consider whether I believe that p independently of considering whether p; and Wittgenstein's diagnosis of the temptation to think this is a familiar one:

'At bottom, when I say "I believe ..." I am describing my own state of mind - but this description is indirectly an assertion of the fact believed'. As in certain circumstances I describe a photograph in order to describe the thing it is a photograph of. But then I must also be able to say that the photograph is a good one. So here too: 'I believe it's raining and my belief is reliable, so I have confidence in it [= so it's raining].' - In that case my belief would be a kind of sense-impression. (my suggested amendment)

To this the reply is:

One can mistrust one's own senses, but not one's belief (ibid. p. 190).

The reason why there is the asymmetry shown by the 'paradox' - why self-ascriptions of belief are 'qualified' avowals in the sense shortly to be summarised - is that they are not 'descriptions of one's state of mind'. 'There is a similarity here to expressions of emotion, of mood, etc.', Wittgenstein writes. Remarks in this section of the Investigations attempt to entice us away from the contrary picture:

I say of someone else 'He seems to believe ...' and other people say it of me. Now, why do I never say it of myself, not even when others rightly say it of me? Do I myself not see and hear myself, then? - That can be said ...

... My own relation to my words is wholly different to other people's. (ibid. pp. 191-2)

There is also the 'Wittgensteinian snort' discussed in another context (see p. 199 below). A 'deep resolution' of Moore's paradox will, however, require an explanation of the authority of belief self-ascriptions.

(vi) What does Evans' account show? Qualified avowals defined

Evans' account, though very similar to Wittgenstein's in its analysis, does not really offer much by way of such an explanation. What does his account show? Let us begin by examining more clearly the notion of 'qualified' avowal. From the discussion of the location of error in (to take our central example so far) belief self-ascriptions, and of Evans' account, we can now say how far these self-ascriptions conform to the Famous Five. They do so, in fact, in a qualified way; the qualifications stem from the behaviour of the propositional element, and the consequent non-immediacy of self-knowledge in the case of belief (cf. pain). Here are the five features, modified so that they constitute features of 'qualified' avowals - viz. propositional attitude self-ascriptions with belief as the paradigm:

- (i) There are no grounds for the utterance 'I believe that p' over and above the grounds for 'p'.
- (ii) IEM
- (iii) My very utterance of 'I believe that p' is a criterion for the correctness of 'He believes that p' said of myself by another - except when the subject has misidentified the object of his belief (see below).
- (iv) When X, comprehendingly and without intending to deceive, utters 'I believe that p', the truth of 'He believes that p' said of X is guaranteed - except as under (iii).
- (v) 'I doubt whether I believe that p' has no use except as an expression of doubt whether p, and no use at all after the utterance of the avowal 'I believe that p'.¹

Features (i), (ii) and (v) are each directly connected with the

¹ Intervention of the Unconscious aside.

Evans principle. Because my beliefs are something I have a part in forming (they are not simply 'given'), because they are 'plastic' and responsive to information, my knowledge of them cannot be 'immediate' as in the case of pains. There may be a period when, since I am considering whether p, there is no fact of the matter about whether I believe it or not. And we have already seen how misleading it is to ask certain questions (about whether my attitude towards p is one of belief or something else) which imply that there is always such a fact of the matter. Features (i) and (v) show that the presence of grounds and the possibility of doubt always pertain to the embedded proposition (the propositional element) and not to the complete self-ascription.

There is a minor difficulty in the suggestion that 'qualified' avowals straightforwardly exhibit features (iii) and (iv) of our original formulation. Suppose I say, comprehendingly and without intending to deceive, 'I believe that Bill Bailey won't be going home tonight', intending to refer to a man slumped in the corner of the crowded room rather the worse for drink. But I have misidentified the object of my belief - it is not Bill Bailey but Fred Smith. It would not, I presume, be correct for someone to say of me 'He believes that Bill Bailey won't be going home tonight'. Assessing the significance of this point involves opening a can of worms; clearly, however, as in the modification of the Famous Five just discussed, the qualification stems from the behaviour of the propositional element alone.¹

It is worth noting at this point that we have arrived at a more rigorous account of what Rorty terms 'mental features' (cf. 'events') than is provided by that author in his interesting article 'Incorrigibility as the Mark of the Mental' ([1970] pp. 399-424). He exaggerates the extent to which avowals of propositional attitude can be subject to honest mistake, presumably because he has not carefully distinguished the merely apparent from the genuine cases (see under 'The propositional element', pp. 152-56), and writes:

¹This example is related to those of de re beliefs discussed under the earlier heading 'The propositional element' (pp. 152-56).

Statements about beliefs, desires, emotions, and intentions ['mental features'] are implicit predictions of future behaviour, predictions which may be falsified. Such justification provides an accepted procedure for over-riding [1st-person] reports. In this they are distinct from reports of thoughts and sensations ['mental events'] which are compatible with any range of future behaviour.

But the fact that we are not incorrigible in our reports of mental features as we are about mental events should not blind us to the fact that we are almost incorrigible. The possibility of over-riding reports about such features is real, but is actualised only rarely ... Further, as such mental features as beliefs and desires become more particular and limited and, thus, approach the status of episodes rather than dispositions, they become more incorrigible (ibid. p. 420).

The initial point to note is the ambiguity in the first sentence quoted - clearly, 3rd-person ascriptions of propositional attitude are implicit predictions, but the way in which 1st-person ascriptions figure as 'qualified' avowals makes it seem highly dubious that one should say the same about them. Further, we have given content to the 'almost incorrigible' status of the latter by describing them as 'qualified' avowals - and discussing the cases where the possibility of over-riding such avowals is actualised. The explanation of this intermediate 'qualified' avowal status in terms of reference to an intermediate and fluctuating position on an episode-disposition continuum looks enticing - but as in all generalisations about the mental, care is needed to avoid being trapped by pictures of appalling crudity. Less tentative judgement awaits discussion of other self-ascriptions - and I consider Rorty's views a little further in my treatment of the central question of expression (pp. 227-29 below).

Now in answer to the question 'what does Evans' account show?', I will say 'It shows that self-ascriptions of belief are "qualified" avowals' in the sense just given; and little more. For Evans has not explained the authority that the 'qualified' avowals express (our central problem again); how is it that the procedure I put into operation for deciding whether p does not, in addition to deciding whether I believe that p, also decide whether my mother

believes that p? It might be thought that Evans has given us something in his observation that 'one's eyes are ... directed outward - upon the world' in belief self-ascriptions. But the Wittgensteinian inspiration for this observation should bring to one's attention the fact that the data Evans adduces (for that is what they are) are equally compatible with a non-assertoric account - that belief self-ascriptions are non-assertoric speech-acts which do not serve to express self-knowledge. In response to Evans' claim that 'if a judging subject applies this procedure [as outlined], then necessarily he will gain knowledge of his own mental states', one may say: 'That is certainly correct, but you still owe us an account of what this knowledge consists in'. Such an account forms the conclusion of this chapter; and the particular case of 'qualified' avowals is treated on pp. 231-33.

4.3) Avowals of Desire

I will now discuss desire as the paradigm of a propositional attitude with a different 'direction of fit' with the world to that of belief, before moving on to perceptual experience and other mental states closer to the 'pain' paradigm. The analogies with belief will become apparent. No more need be said on the first two possible sources of error in the self-ascription 'I desire that p'; clearly it must be IEM, and the possibility of error in the attitude component has already been discussed. In connection with the propositional element (the embedded proposition), the points made about belief will carry over to desire; but the analogies with errors in the 'normal' (non-unconscious) cases need spelling out.

As in the case of belief, some apparently mistaken avowals are not genuinely so. Take generally debilitated Gerald, who avows 'What I want is a change of job'. So he gives up his employment as a highly-paid computing engineer and goes off to manage an hotel in the Outer Hebrides - where he is thoroughly miserable, missing the intellectual challenge of computing. One should not too readily assume that Gerald was mistaken in his avowal. More likely (though it would depend on a fuller description of the case) one should say that he made a mistake about the consequences of satisfying it; that if he'd consulted his doctor he would have discovered that his debility was in fact the result of a thyroid condition or, more picturesquely, a mid-life crisis, which a change of job would be ineffective in curing or resolving. In this the example would be analogous with one option in the example 'I believe that socialism is the best system' discussed earlier - in each case the subject comes to realise he has been mistaken about the consequences of the embedded proposition (or rather, the causal consequences of its obtaining).¹

¹ Likewise analogously with belief, 'I'm not sure whether that's what I want' is not an expression of ignorance about one's determinate mental state, but rather a reflection of the

More so than in the Thatcher example of allegedly mistaken belief, the pressure to regard the subject's avowal of his desire as mistaken may persist. It is well to understand the origin of this pressure. Part of the trouble may lie in an equivocation in the sense of 'satisfaction' when one talks of the 'satisfaction' of a desire. It is important to note that the satisfaction of a desire need not be satisfying; if a desire is satisfied, I need get no satisfaction out of it. However, we generally look for an explanation of the latter (unusual) phenomenon by constructing a practical inference schema. In order for us to regard a given case as one where the satisfaction of a desire has been unsatisfying, rather than saying that the subject made a mistaken avowal of the desire in the first place, we need to place the desire in question in the context of such a schema. It is then regarded as the desire satisfaction of which is incorrectly believed to be the means towards satisfying some higher-level desire. Thus Gerald wants to slough off his debility, and his desire to change his job is the unfortunate product of reasoning that a change of job is the means to the former end. (Questions about the justificatory compared with the explanatory role of the practical syllogism, and about how the latter can be fulfilled where it appears no conscious deliberation has occurred, cannot be allowed to detain us here.) This practice suggests that there are two principles about desire that need to be respected, tension between which the practice serves to lessen:

- (A) The desires ascribed to an agent must generally be such that their satisfaction is satisfying to him.
- (B) An agent's sincere avowals of desire are generally correct.

It is interesting to note how principle (A) can begin that process of divorcing desire-ascription from avowal which has as its end-result the justification of political tyranny. Both Rousseau's 'forced to be free' dictum and Marxist accounts of the 'false consciousness' of those who adhere to 'bourgeois freedom' may make implicit appeal to the idea that an agent only really wants what is (according to Rousseau or Marx) in his fundamental interest. The

indeterminacy of that state at that time (cf. Wittgenstein: Z 57).

authority of avowals of desire is, on such accounts, a heavily qualified one - if their talk of desire is other than a façon de parler. For present purposes, it is important simply to note that constraints on principle (A) are supplied by the need to respect principle (B).

Cases of genuine error are, as in the case of belief, two - cases where the Unconscious is implicated (on which no more need be said), and cases where the object is misidentified. Say I experience the delights of a liqueur known as 'Burns' Scotch Cream', but subsequently confuse it with the more well-known 'Bailey's Irish Cream'. I then avow 'I desire a measure of that excellent Bailey's Irish Cream' - a liqueur which I have never in fact tasted. It is clear in this case that what I in fact desire is Burns' Scotch Cream - as becomes apparent when I am offered some Bailey's, find it to be disgusting and cast around for an explanation. The importance of causal origin in determining the direction of a desire upon a particular object is apparent here (analogously to its role in the case of thought and object - see e.g. Wright [1983] pp. 100-101).

There is a certain picture of desire which makes misidentification seem a more common possibility than it is in the case of belief. Desires can be seen as forces to which propositional contents are attached but which are primarily to be identified in terms of their 'fate' rather than their object.¹ One may indeed not be conscious of desires at the appetitive or impulsive end of the desire continuum (as opposed to desires which are the product of practical reasoning). The self-ascriptions of such desires may turn out to be non-avowals ascribed on a behavioural basis rather than 'qualified' avowals.² But in the case of the kinds of desire the self-ascription of which is authoritative, the Evans principle

¹ Such a picture is favoured by Freudian theory, with its requirement of persisting desires with transient objects (see Wollheim [1979] p. 47).

² If the use of 'desire' in question is extensional, the apparent avowal will be non-authoritative. Such cases will be analogous to 'I enjoy ...', which is extensional.

about the direction of the subject's attention carries over from belief. If I'm wondering whether I believe that p, I'm not doing anything over and above wondering whether p is true. Likewise, if I'm wondering whether I desire that p, I'm not doing anything over and above wondering whether p is good (pleasurable, useful or whatever). (R.M. Hare, in his [1952], attempted to demonstrate a tight connection between goodness and preference, but his view has been disputed, and it would seem unwise to insist on a similarly tight connection between goodness and desire.)

4.4) Avowals of Perceptual Experience

(i) 'Sense-data' and non-avowals of perceptual experience

The question of the self-ascription of perceptual experience is of such difficulty it makes that of the self-ascription of propositional attitudes seem quite straightforward in comparison. Once again we will turn to Evans for assistance, but will find that his insights are possibly mixed with confusions.

Self-ascriptions of perceptual experience enjoy a significant position in epistemology, because of their historical role in attempts to discover alleged foundations of empirical knowledge. The recourse frequently made to them as expressions of certain knowledge on which the rest of our epistemic commerce with the world is based is a symptom of their status as avowals. In the empiricist tradition, avowals like 'I seem to see something red', '... hear a high-pitched noise', '... feel something hot', refer to 'sense-data' or 'sense-impressions' with which the subject's acquaintance is direct, immediate and certain. Claims about the world ('I see something red', etc., being minimal instances) may then be 'inferred' from these avowals; the 'sense-data' thus have a logical role as well as the causal role they play as mental products of the physical objects, inferences to the existence of which the avowals license. This dual role is worth bearing in mind in the context of discussion of more sophisticated accounts of perception - including that of Gareth Evans, whose 'information-states' may trace their ancestry, in some distant way, to sense-data.

Unfortunately for the sense-datum theorist (though this may be the least of his troubles) the class of avowals of perceptual experience is both more limited, and also perhaps (more seriously) unacceptably wider, than he may want to say. Firstly, 'appearances' can be 'objective' too. For it is of course possible to imagine a dispute over whether something looks red; think of two parties looking at cave drawings in poor lighting conditions, for instance.

'It looks red' isn't grammatically an avowal, of course; but as J.L. Austin notes, even when the claim becomes 1st-personal and thus 'safer', it need not thereby be made incorrigible. Adding an 'It seems to me that' or an 'I seem to see' operator will not automatically convert the claim in question into an avowal in our sense of the term. Austin, in the course of his remorseless and conclusive demolition of sense-datum theories ([1962]), points out that I can be honestly mistaken in the self-ascription of perceptual experience involving any but the very simplest observational concepts - and perhaps even with the latter:

'It seems to me personally, here and now, as if I were seeing something magenta' ... may be more cautious [than 'That is magenta'], but it isn't incorrigible ... there is always the possibility not only that I may be brought to admit that magenta wasn't the right word to pick on for the colour before me, but also that I may be brought to see ... that the colour before me just wasn't magenta (ibid. p. 113).¹

Evans similarly contends that 'I seem to see eleven points of light arranged in a circle' is not immune to honest mistake - the subject may have miscounted, forgetting where he began. He goes on to comment that in the case of very simple concepts (colour words like 'red') it is hard to make sense of a mistake:

[but] this sort of infallibility is rather limited and uninteresting. And it is of a quite different kind from that which arises in the case of the self-ascription of belief (ibid. p. 229).

(We shall consider these comments later - see p. 175.) In conclusion, then, claims about what I seem to experience, about what things appear or look to me to be like (a mixed collection) do not always constitute a retreat into personal experience on which the subject has authority. Rather than viewing such self-ascriptions as making definite claims about a percept, rather one should view them as making tentative (though still revisable) claims about an object -

¹ But see p. 173 below.

or sometimes as making no (further) claim at all. If, to take an Austinian example, I am challenged in my claim 'I see a pig', I may reply 'It looks like a pig' - another statement about the world. Or perhaps 'It seems to me like I see a pig' - more a retreat to one's disposition to make the original claim than an expression of or report on personal experience.

(ii) 'Avowal' is a category of tokens, not types

This brings us to our second point. For just as 'IEM' is applicable to a category of sentence-tokens rather than sentence-types, so is 'avowal'. This point should have been apparent from the comments on 'non-avowal' uses of e.g. 'I am afraid'; which is not to say that all avowals have corresponding non-avowal uses (e.g. 'I feel hot'). It is thus apparent that, under certain conditions, plain statements of the form 'I see ...', 'I hear ...' etc. may also constitute avowals (to the discomfiture, clearly, of our historical stalking-horse, the sense-datum theorist). However, in order to make out these claims, a non-circular characterisation of the conditions under which some token is an avowal is required. This is so at least for the reason that feature (iv) of the Famous Five already states conditions under which the utterance of 'I am in pain' guarantees the truth of 'He is in pain' said of me; conditions which will now require supplementation, but which already in fact imply that 'avowal' is a category of tokens. The conditions are, if the reader recalls, that the utterance is (a) comprehending, and (b) not intended to deceive. If we want to preserve the notion of an interesting class of avowals, we must be careful to show how these and further conditions do not constitute ad hoc revisions of what is (it will be suggested) a theoretical claim (that there is such a class), made in order to accommodate genuine counter-examples. (Some may think this move has already been made, and that the possibility of error due to thinking that an ache is a species of pain, shows that 'I am in pain' is not interestingly incorrigible.)

The waters are becoming murky. But we can at least say more about what these aforementioned supplementary conditions appear to be! Austin, in the course of his criticism of the notion of

'intrinsic incorrigibility'¹, notes that whether a statement is incorrigible

is not a matter of what kind of sentence I use in making my statement, but of what the circumstances are in which I make it. If I carefully scrutinise some patch of colour in my visual field, take careful note of it, know English well, and pay scrupulous attention to what I am saying, I may say 'It seems to me now as if I were seeing something pink'; and nothing whatever could be produced as showing that I had made a mistake (ibid. p. 114).

The further condition for avowals of perceptual experience which the passage suggests is:

(c) the utterer exercises appropriate care in his observation and judgement.

But is 'appropriate care' strong enough? The formulation

(c)₁ the utterer does not make an avoidable mistake is stronger, but clearly more open to the charge of circularity. Ginet, in the course of his discussion of the infallible justification of knowledge-claims about one's current conscious state, makes a distinction between 'avoidable carelessness or inattention' and being 'the unwitting victim of unnoticed slips in calculation or delusory impressions of sense or memory or misleading appearances or evidence' ([1975] p. 59). When one arrives at this level of refinement one knows one must be approaching the dizzy heights of the Grice-Strawson battle on non-natural meaning; and similar doubts arise as to whether it is the Homeric Opposition or the War of Jenkins'² Ear that is being decided.

¹ Developed e.g. by G. Pitcher in his [1971], pp. 20ff.

² In fact, of course, this conflict between England and Spain, provoked in 1739 when an English sea-captain lost his ear apparently at Spanish hands, was a serious one - developing into the Seven Years' War.

(iii) Is there 'an interesting class of avowals?'

It is worth noting that the range of observational judgements, tokens of which can constitute avowals, needs to be limited in a further way, moreover. For presumably the concepts involved must be non-technical; otherwise the requirement of comprehending utterance will just be an ad hoc sponge to soak up counter-examples. Thus, Austin's earlier 'magenta' example would not in fact constitute an avowal, whereas 'I seem to see something pink' would. (Many people would need a colour chart to decide whether something was magenta; the use of such a criterion would violate feature (i) of the Famous Five, at least.) But Austin would not be sympathetic to the idea that there is an 'interesting class of avowals'; there are just lots of different kinds of sentences which may be uttered in making statements that are in fact incorrigible (ibid. p. 115). He wants to say that there is no interesting difference in respect of incorrigibility between 'I seem to see something pink' and examples such as the following:

... if I watch for some time an animal a few feet in front of me in a good light, if I prod it perhaps, sniff, and take note of the noises it makes, I may say, 'That's a pig'; and this too will be 'incorrigible'; nothing could be produced that would show that I had made a mistake. (ibid. p. 114)

It is clearly necessary for an avowal to have 1st-personal grammatical form; and 'That's a pig' is ruled out because it is not an avowal of some state of the subject, and hence cannot exhibit features (iii) and (iv) of the Famous Five which express the subject's authority. It might be thought, however, that the similarities between the two kinds of example Austin adduces at this point are sufficiently close, and the differences such grammatical inessentials, as to show that:

- (1) There may be no interesting class of avowals
- and/or (2) The solution of the 'problem' of avowals lies in showing how the incorrigibility of the latter is nothing unique or special.

But this is mistaken. Our problem with the self-ascription of

perception arises just on the assumption that there is an interesting class of avowals the conditions on which we do not want to over-dilute; so cannot constitute an argument against the proposition that there is such a class. What marks out the class of avowals as interesting is that the subject cannot honestly be mistaken about the state he is in, and so feature (iv) of the Famous Five specifies as conditions only sincere and comprehending use.

However, it might be insisted that perceptual self-ascriptions have to be central to the class of avowals, without which the claim of an interesting, unified such class lacks credibility. The following argument might then be pressed:

Your claim dies the death of a thousand qualifications. I present a counter-example to your contention that a certain range of utterance-tokens exhibit the incorrigibility of avowals, and you promptly add on an extra condition accordingly. And so on. But your conditions are either insufficient or circular; which goes to show that they are mere ad hoc responses to a contention which is surely correct, viz. that one can specify no genuine unifying features common to this so-called class of avowals.

Austin would be sympathetic to the conclusion of this line of reasoning, if not to its premises. It is more likely to be deployed by a physicalist for whom the thought that the results of a scientific psychology must be beholden to a dogma of (obnoxious phrase) 'folk-psychology' (viz. the authority of avowals) is intolerable. Indeed, he would do better to deploy it than give time to some of the more tendentious objections to this dogma.¹

Now, it is certainly true that the Famous Five are meant to be

¹ See e.g. Armstrong [1963] p. 424. He supposes that a correlation indicating identity has been established between a sensation of red and a brain-state - say 'no. 143'. It is conceivable, he suggests, that the subject should truthfully report such a sensation when he is not in that brain-state - under laboratory conditions including administration of a

'salient features' of the class of avowals, and not necessary and sufficient conditions. Can one not then reply to the critic that the fact that the conditions are 'either insufficient or circular' doesn't matter? (One might go on to contend that avowals exhibit a mere 'family resemblance'.) Well, it is certainly true that we're not bothered about the kind of insufficiency implied by the following example. An actor on stage says, comprehendingly and without intending to deceive (and in all seriousness - acting is a serious business), 'I am in pain'. Yet the utterance is not, in contrast to other sincere and comprehending uses, an avowal. This is no worry to the present account, because no one would ever take such an utterance to be an avowal. (Indeed, mention has earlier been made of non-avowal tokens of e.g. 'I am afraid', etc. - though not, of course, of 'I am in pain'.) And we are not trying to build up an exhaustive, sufficient account that will exclude cases like that. However, the kind of insufficiency apparent in the perceptual cases is more important. Here, something looks like an avowal, but then fails to yield the appropriate guarantee as in Feature (iv) of the Five. Our account seems to be insufficient because it fails to include such cases.

The objection of 'death by a thousand qualifications' overstates the case however. Perceptual self-ascriptions are important, but it is not essential that they be incorporated in the category of 'paradigm' avowals. Given this fact, it would be preferable to retain the integrity of the latter as a category where 'honest mistake' is impossible, and to view 'avowals of perception' as akin to 'qualified' avowals, though puzzlingly more open to honest error. This would not be to endorse Evans' over-confident view that the infallibility in perceptual self-ascription is 'rather limited and uninteresting', and quite different from that of the self-ascription of belief.

'truth-drug'. Armstrong's unquenchable scientism leads him to conclude that it is therefore possible for someone to believe he has a sensation of red when he doesn't have it.

(iv) Evans on self-ascription of perceptual experience;
'intermediate avowals'

We are left, then, with the conclusion that the avowal-status of perceptual experience is at best harder to define than that of self-ascriptions of belief; paradoxically, it might seem, given that the former are closer to the sensation paradigm of avowal than are the latter. Be that as it may, there is the same temptation in both cases to construe the self-knowledge involved (if such it is) on the model of 'looking within'. It would be useful, therefore, to examine an account of the self-ascription of perceptual experience (that of Evans) which, for all that it fails to discuss the kind of complexities in that experience brought out by our preceding discussion, does aim to avoid such a model.

A perceptual experience is, according to Evans, an 'informational state' of the subject. In the present connection, such states share with sense-data what appears to be both a causal and a logical role (or some mixture of the two). They are 'internal', possess 'content' ('the world is represented in a certain way') and some of them serve as 'input' to the 'concept-exercising and reasoning system'. Judgements are 'based upon (reliably caused by)' these internal states; but though they possess (non-propositional??) content, the states are 'non-conceptualised'. The crucial point which Evans insists on is that though the subject's perceptual judgements are 'based on informational states', they are not 'about' those states:

His internal state cannot in any sense become an object to him.
 (He is in it.) (ibid. p. 227).

A subject can gain knowledge of such states (that is, I presume, can come to know simply how things appear to him to be) - but only in a roundabout way, by

trying to make a judgement about how it is at this place now,
 but excluding any knowledge he has of an extraneous kind¹
 (ibid. p. 227).

¹ Evans refers us to Dummett's 'What is a Theory of

In contrast to the case of belief, though, this procedure does not produce infallible knowledge of the mental state - as we have already seen from the '11 points of light' and other examples (see p. 170).

Now, Evans is quite explicit in denying that an informational state can have a justificatory or inferential role with regard to some perceptual judgement. That is, I do not and cannot (as sense-datum theorists would require) infer, from the fact that I have some informational state, that the world is or appears to be thus-and-so; nor can I justify the latter judgement by appeal to my informational state. Such states are not, unlike the sense-data with which they are being compared, objects to their subject. However, it seems to the present writer that, from what it is possible to understand of Evans' highly compressed and abstract account, one ought to have some misgivings on this score. He comments that, in contrast to the case of belief, with perception

there is something (namely an internal, informational state of the subject), distinct from his judgement, to which his judgement aims to be faithful (ibid. p. 230).

Though Evans goes on to insist that this 'something' is necessarily to be approached in the roundabout way he has described (by trying to make a judgement about how it is at this place now, but excluding any extraneous knowledge), one wonders how nonetheless a judgement can aim to be 'faithful' to something that is not an object to the judger. Further doubts arise from the statement that judgements are 'based on (reliably caused by)' the informational states. Underlining of the relation does not help to make it clearer; the suggestion that only non-deviant causal chains can be allowed

Meaning? (II)' ([1976] p. 95) for an explanation of 'extraneous'; but the single sentence there (a short one by Dummett's standards) is not greatly illuminating. The idea seems, however, to be that inferential knowledge is ruled out; e.g. if I seem to see a dagger, but know I have taken a hallucinogen, the knowledge I may have that the dagger is delusory counts as extraneous.

between the informational state and the judgement, makes an opponent of causal theories in the philosophy of mind wonder whether the relation between the two might not in fact be a logical one.

Whatever the status of avowals of perceptual experience, there is nonetheless an intermediate sub-category of non-paradigm avowals which yet do not express propositional attitudes. This important sub-category is in fact rather a rag-bag. Avowals of perceptual experience (if such they are) have no connected natural behavioural expression; but nor are they, in a range of cases at least, subject to holistic constraint. Avowals of certain moods or feelings equally do not have connected natural expression, but are subject to holistic constraint. Moods and feelings thus fall into two categories:

- (a) Those which figure in 'paradigm' avowals. These are mental episodes whose origin is best viewed in causal terms, and for which if distressing to the subject the best injunction is 'pull yourself together'.
- (b) Those which figure in the sub-category which, for want of a better term, we may call 'intermediate' avowals. These are more enduring or recurring mental items whose origin should be viewed in intentional terms, and which if troublesome may call for some kind of analysis.

It is important, in everyday life, to be clear about this distinction; much distress can be caused by mistaking the latter for the former, and much time wasted by mistaking the former for the latter. For instance, a feeling of remorse will fit into category (b). It has no natural behavioural expression over and above that perhaps associated with depression - than which it is clearly more refined. Its self-ascription must fit into a coherent self-image the subject presents to the world - in contrast to a feeling of depression, the relative transience and non-intentional provenance of which mean that it need not so fit.¹ Holistic constraints notwithstanding,

¹ A feeling of remorse must be accompanied by a number of beliefs about one's responsibility, the unhappy nature of the

we are peculiarly dependent for our discernment of intermediately avowable states on the subject's verbal expression of them.

state of affairs one has brought about; and by a desire to ameliorate that state of affairs, perhaps. One might then say, that although 'intermediate' avowals do not express propositional attitudes, they imply them.

4.5) An Argument to the Non-Assertoric Thesis(i) Statement of the argument from non-justifiability

We now turn from discussion of the different sorts of avowal to the general issue. In my account of the arguments from IEM to the 'no-reference' view, I contrasted the conclusion there with that of the arguments for the non-assertoric thesis, suggesting that in the latter case there was a motive for accepting it, but in the former case there were simply the arguments (p. 101). This was slightly unfair on the 'no-reference' view; for since arguments for the non-assertoric thesis, as distinct from mere anti-Cartesian motives for its acceptance, have been rather thin on the ground, one might in fact find the contrast favourable to the former. Is the non-assertoric thesis a mere proposed panacea, a label for the desired solution?

It should come as no surprise that given the Wittgensteinian pedigree of both positions, some useful parallels may be drawn in the effort to give an argumentative justification for the non-assertoric thesis. Here is an analogue of principle (K) for knowledge (p. 104 above):

(A) 'X sincerely asserts that p' is meaningfully assertable only if possibly on the same occasion (p and X sincerely asserts that ¹ $\neg p$).

Why should one adhere to such a principle which has the effect of legislating avowals as non-assertoric? In the case of (K), we relied on the idea that in cognitive as well as other kinds of achievement, failure must be a possibility for the notion of achievement to be applicable. How does 'sincere' assertion involve

¹The 'same occasion' qualification is necessary even in its possible vagueness to accommodate the fact that 'avowal' is a token-category.

a cognitive achievement? Not in a direct sense. There is some requirement on the speaker's comprehension of the expressions used in the utterance, which would require a cognitive achievement (language-learning). But the 'indirect' connection I have in mind between assertion and cognitive achievement is rather one derived from arguments of Huw Price ([1983] pp. 354-7).

Price is concerned inter alia to define the notion of assertoric force, and commences by discussing one function of the terms 'true' and 'false' - 'to provide a uniform means of endorsing or rejecting a statement made by a previous speaker' (ibid. p. 355).¹ This characterisation is insufficient as it stands, Price argues, though a step in the right direction is made by Dummett when he writes

the roots of the notions of truth and falsity lie in the distinction between a speaker's being, objectively, right or wrong in what he says when he makes an assertion ([1978] p. xvii).

We need the notions of 'right' and 'wrong' here because questions, commands, requests etc. can be 'endorsed' or 'rejected' just as assertions can; yet only assertions can be true or false. Yet we still need an account of the function and consequences of the forms of criticism ('right', 'wrong', 'correct', 'incorrect', etc.) that Dummett has in mind. Price writes:

It seems ... that the primary significance of these forms of criticism lies in the fact that they constitute a challenge to a speaker to justify an utterance, and an indication of readiness on the part of the critic to engage in a dispute ... Ideally the more well-justified [view] prevails ... Plausibly, there is enough of a general advantage in such dispute behaviour to explain the existence of a powerful linguistic device to

¹ He also genuflects to Davidson, being concerned to raise a difficulty for radical interpretation and the notion of 'holding true'. But what he has to say has an importance exceeding such merely fashionable concerns.

facilitate it (i.e. the use of 'true' and 'false') (ibid. p. 356).

When one looks at society today, the prevalence of better-justified views must seem very 'ideal' indeed; yet such fin de siècle pessimism does not affect the philosophical point. Price goes on to sum up his argument by substituting 'belief' for the imprecise 'view'; disputes between speakers over utterances expressing, say, desires would seem to have little point, whereas

some utterances (call them 'assertions') characteristically express states of mind ('beliefs') with respect to which there is reason to seek agreement between speakers (viz. the 'general advantage' above) (ibid. p. 356).

Where does this connection of assertion with justification get us? Well, if truth or falsity are predicable of any assertion, and only of assertions, and if the function of such predication is as outlined, then an assertion must be something for which it is possible to give a justification.¹ But if it is not possible on some occasion both for p to be true and for X sincerely to assert its negation, then on that occasion the idea of a justification for the putative assertion that p has no place. When Mrs. Thatcher makes some claim about the principles of management of the economy, say 'It's just like running a very large grocer's shop in Grantham', she may think that justification for such a self-evident proposition has no place, but nonetheless it is possible.² ('Justification'

¹ This may amount to the idea that to assert something is to make a knowledge-claim.

² The Thatcher example may in fact indicate that the principle should be stronger; not just that there can be a justification, but that on the occasion in question the subject is prepared to offer one. Indeed this strengthening may be necessary in order to anchor 'assertion' as a token-category; since the attitude of the subject towards the utterance will be part of the 'occasion of utterance'. What I have in mind is this. If Mrs. Thatcher is not prepared to offer a justification for her putative assertion, it might be

does not mean 'conclusive justification'; simply that reasons can be given or evidence adduced in support of the proposition.) As in the case of knowledge, principle (A) expresses the requirement that a mistake must be possible for something to count as an assertion - in this case, further, since we're concerned with a linguistic act where deceit is possible, it must be possible for the subject to be honestly mistaken. (Hence the qualification 'sincere assertion'.) The rationale for this requirement in the case of knowledge stems straightforwardly from the status of knowing as an achievement. In the case of assertion, the connection is with the 'achievement' of justification via an account which marks off assertoric from other speech-acts in terms of the unique appropriateness of dispute-behaviour and its linguistic facilitation, which would be otiose if the kind of mistake specified in (A) were impossible. Avowals will fail test (A), and therefore, so the argument goes, cannot be assertions.

(ii) Responses

It may be thought that the connection of assertion and justification has its origin in an idea that renders dubious the application of the connection to the case of avowals. For surely, it might be objected, the rationale for making the possibility of justification a requirement for assertion consists in the need to rule out pseudo-assertion; a disreputable kind of speech act that succeeds in saying or expressing nothing. Take the following conversation overheard recently in a London pub, let's say:

A: I've been to the moon, you know.

B: Get away, when was that?

more happily viewed as a 'framework principle' in the sense of Wittgenstein's On Certainty and thus possibly not as an assertion. This function might then explain the proponent's unwillingness to abandon the principle; this in the face of the significant body of evidence pointing to the catastrophic national effects of the wholesale rejection of Keynesianism which it implies. (This would be a charitable explanation.)

A: Oh, sometime ...

B: How did you get there? One of them space-shuttles, was it?

A: Oh no.

B: A rocket then?

A: No ...

B: How did you get there then?

A: Oh, I just ... went ...

B: Love-a-duck! What was it like then?

A: Hmm ... so-so ...

¹
Etc.

A puzzling encounter indeed. It seems that A does not recognise the need for some justification for his opening remark; suggesting that it is a vacuous pseudo-assertion rather than, say, a lie. (Whether it is a piece of mischief-making or an altogether stranger phenomenon seems immaterial.) The phenomenon objected to is the kind of unsavoury item intended by the Logical Positivists to be impaled on the horns of the dilemma 'false or meaningless'. In contrast, it may be argued, avowals (or putative avowals) are rarely in either of these latter categories, since, when sincerely uttered, they are true (cases of uncomprehending utterance being quite bizarre). The absence of justification is a bar merely to the assertoric status of utterances one would not want to 'endorse'.

A different response would be to say not that the unjustifiability of avowals is of the sort compatible with assertoric status, but rather that they are not, in the relevant sense, 'unjustifiable'. For it is possible for a speaker to justify his avowal, not in the sense of showing that it is not the product of a mistake, but in showing that it is sincere. The 'dispute-behaviour' Price talks of would thus take a different form. It is true that avowals are treated as sincere unless proven otherwise (see pp. 224-25 below) so such disputes could not be common. But there is in any case always the possibility of a 3rd-personal justification of X's avowal to a third party, and of a dispute there. (E.g. X avows

¹
I am indebted to Michael Luntley for the idea of such a conversation - suggested to him by the passages in On Certainty on lunar travel (e.g. para. 106).

he is full of joy, and two observers subsequently disagree over whether X's circumstances are such that one should regard the avowal as sincere.)

The argument from non-justifiability can certainly be criticised, then, in the preceding way and in others,¹ but it should not be derided for what it could not do. The Cartesian, for instance, may well have a different account of assertion to that outlined, to go with his different account of the mental; the fact that, to adapt Wittgenstein's metaphor, we have in connecting assertion and justification put two books together on the library shelf, does not mean that someone else will not feel like carting the whole row off to the bonfire (BB pp. 44-5). We will at best have succeeded in presenting a plausible and coherent picture, and shown what wider view the assertoric proponent will be committed to endorsing or rejecting. (Though the account of assertion implies principle (A), the converse does not hold; but it is difficult to conceive of an alternative, and unconnected, rationale for the principle.)

(iii) The 'anti-Cartesian motive' for the thesis

Moreover, deciding the correctness or otherwise of the Price-inspired argument does not seem to be crucial for present purposes. This is because although it has the non-assertoric status of avowals (but also of many other putative assertions) as a conclusion, it does not seem to have any specific anti-Cartesian motivation of the sort connected with our central problem. The real argument for the non-assertoric thesis, though it will not be analysed as such, is that labelled earlier as a 'mere anti-Cartesian motive' (p. 180 above).

This argument can take various forms; like the arguments from IEM to 'no-reference', it tends to be implied in the literature and

¹ For instance, one would have to consider the status of other apparent assertions ruled as non-assertoric by principle (A); e.g. analytic truths and indexical non-avowals that are incorrigible.

not spelled out (see p. 100). One version mentioned at the start of this chapter (pp. 145-47) goes like this:

(1) If avowals are assertions, they must have truth- or assertion-conditions.

(2) Avowals are assertions.

So (3) They have truth- or assertion-conditions.

(4) Such a condition must be the obtaining either of an 'inner' or an 'outer' state of the subject (one to which he has an access of the sort denied to others, or one to which he has the same kind of access as do others).

(5) The condition cannot be 'outer' (see p. 146).

So (6) It must be 'inner'.

So (7) The subject's authority expressed by his avowal must be based on his special access to an inner state on the model of expert knowledge.

But (8) This account of authority involves Cartesian incoherence.

So (9) (~ 3) Avowals do not have truth- or assertion-conditions.

So (10) (~ 2) Avowals are not assertions.

As must be apparent, this is a rather 'impressionistic' presentation. It is simply meant to reflect the sort of thinking that is going on when people propose a non-assertoric thesis. The argument will not be analysed further here; but it will become clear in the course of discussion that it is the move from (6) to (7) (at least) that is questionable. We need first to see what the conclusion of the argument amounts to, however.

4.6) 'Expression' (I)(i) The intuitive notion

If avowals are not genuine assertions, what force could they possess? For a start, we return to the idea first canvassed on p. 143. There, the important remark from Zettel on the expressive (cf. informational) role of sentences in the 1st-person present tense was quoted. We need now to explore sympathetically the idea of 'expression' which Wittgenstein, in his few cryptic remarks on the topic, seemed to hope offered a way forward out of the mire. I will begin with a few broad brushstrokes - an account of the intuitive notion of expression, and (in the next section) of the relation of the expressive thesis to behaviourism. We will then need to return to refine the notion of expression along lines suggested by Charles Taylor's article [1979].

The first point to note is that just because an utterance is 'expressive' in an ordinary sense of the term, does not on the face of it mean that it cannot be an assertion (or some other kind of speech-act). In a certain context we can say of each of the following assertions what they express:

- (i) 'Of course, you wouldn't understand that ...': assertion expressing contempt.
- (ii) 'Come up and see me sometime' (Mae West): request expressing interest.
- (iii) 'Don't leave me!': request expressing anxiety.
- (iv) 'Go, and take your wretched dog with you!': command expressing disgust.

The difference between such examples of expression, and the kind of expression involved with avowals, is not quite that it is only particular tokens of these utterances which express the attitude in

question. For it is only particular tokens of e.g. 'I am full of joy' which express joy - i.e. there are non-expressive uses, though every use stands in some relation to the subject's joy. (Or so one would put it as a corrective to an imprecise diagnosis.) It is rather that there is a logical relation between the avowal-type and the mental state in question, such that characteristically utterance-tokens express that state.¹ 'Of course, you wouldn't understand that' is only contingently related to the attitude of contempt; it is not part of its meaning that it expresses such an attitude, and must be said in a certain tone of voice and with a certain facial expression in order to do so.² Said with a different tone of voice, it could express a patronising attitude or (with a smile as well) an avuncular one, perhaps. In contrast, 'I believe that p' is logically related to the subject's propositional attitude; it characteristically expresses (rather than, say, reports) that attitude, and (perhaps consequently) is a criterion for it. In saying 'I believe that ...' a subject may be expressing any number of things, but one could not be said to have understood his utterance at all unless one knew that he was expressing his belief, and that his utterance was a criterion for his having that belief.

This meaning-requirement will be an essential feature of the putative class of merely expressive speech-acts. That it is not sufficient to characterise that class is brought out by considering the following examples:

(v) 'What he achieves is something the rest of us can scarcely even aspire to': assertion expressing admiration.

(vi) 'He is a wicked fellow': assertion expressing moral disapproval.

¹ I mean these locutions to be taken in an anodyne way; Wittgenstein would presumably have said that the decisive move in the conjuring trick had already been made, however (PI 308).

² In contrast, avowals are made in a characteristic tone (see p. 226).

One would not understand such sentences if one did not realise that they characteristically (on the limited occasions of their utterance) express the attitudes in question. These and utterances like them are criteria for these attitudes, it would seem; but we need to be careful in saying this. For maybe 'Go, and take your wretched dog with you!' and utterances like it are criteria for disgust. And so on. So perhaps our initial demarcation within expressive utterances of those which are, and those which are not, logically related to that which they express, does not get us very far? Well, there are clearly mere symptoms of e.g. admiration; in the case (v), the speaker might want to carry his hero's books, hold doors open for him, etc. Expressions of disgust as in example (iv) might occupy a grey area between symptoms and criteria. This does not imply that the class of expressive utterances that constitute some of the criteria for a mental state is quite indeterminate and that the idea of an avowal does not centrally involve a logical relation between an utterance and that which it expresses.

We can add to this account, however. For the feature of avowal which we have been discussing is expressed in feature (iv) of the Famous Five. One might term this also the reflexivity condition; the very utterance of 'I am \emptyset ' is itself a criterion for the speaker's being \emptyset . The other expressive utterances we have been discussing do not 'mirror the fact' which they express in this way; it is the reflexive way in which avowals operate as criteria for mental states which marks them out therefore.

(ii) Wittgenstein's account

What other features will merely expressive speech-acts possess? In hope of some elucidation we will turn at last to the remarks of the chief proponent of an expressive account of avowals, and pursue the course outlined on p. 187. The difficulty of Wittgenstein's account is of course notorious; but then what would be the point of entertaining a less than notoriously difficult account of something as familiar and difficult as the notion of avowal? The idea of a 'solution' to the problem of avowals is in a certain sense to be derided - perhaps the best that can be hoped is that it will come to seem less inevitable that the alternative to

Cartesianism must be a sophisticated form of the behaviourism which Wittgenstein was also at pains to repudiate.

The most well-known illustration of Wittgenstein's expressive account is his suggestion that self-ascriptions of pain supplant natural, unlearnt pain-behaviour. He enquires how a human being learns the meaning of 'pain':

Here is one possibility: words are connected with the primitive, the natural, expressions of the sensation and used in their place (PI 244).

That this tentative suggestion will not do in the case of other avowals is brought out by a remark in Zettel however:

Pain differentiated from other sensations by a characteristic expression (Z 483).¹

Connected with this suggestion, though, is the idea that avowals of e.g. pain might be involuntary and non-intentional linguistic behaviour. There is an ambiguity in some of the formulations here, and we must sort this out before we ascribe to Wittgenstein this interpretation of 'expression'. For instance, Evans talks of Wittgenstein encouraging us to view avowals as 'unstructured responses to situations' ([1982] p. 209n). This might seem to suggest that it is explanation merely at a stimulus-response level that is applicable to the 1st-person psychological utterances in question; that the notion of the subject controlling or intending to make the utterance is misplaced. It is clear, however, that such an austere picture hardly fits more than a limited range of the class of avowals. A more sympathetic interpretation of Evans' phrase is indicated, however, when one considers various remarks in the Philosophical Investigations and Zettel which suggest that avowals are 'reactions' or 'responses' to the mental state (not to an (external) situation, as Evans writes). Expectation and intention are two examples:

¹ Though emotions commonly do have 'characteristic expression-behaviour' associated with them (Z 488).

The statement "I am expecting a bang at any moment" is an expression of expectation. This verbal reaction is the movement of the pointer, which shows the object of expectation (Z 53).

I reveal to him (the listener) something of myself when I tell him what I was going to do - Not, however, on grounds of self-observation, but by way of a response (it might also be called an intuition) (PI 659).

But is it not peculiar that there is such a thing as this reaction, this confession of intention? Is it not an extremely remarkable instrument of language? (Z 39)

Memory is another example:

The words with which I express my memory are my memory-reaction (PI 343).

The equation of confessing and 'reacting to' one's intention suggests behaviour of a different order to that susceptible to a stimulus-response model.¹

However, the latter does seem to be invoked in another context:

... think of the sensations produced by physically shuddering: the words "It makes me shiver" are themselves such a shivering reaction; and if I hear and feel them as I utter them, this belongs among the rest of those sensations (PI p. 174).

There are indeed difficulties in pinning down just what an expressive account is, since Wittgenstein is concerned to point out the range of different kinds of utterances constituted by avowals - from the crude response to a stimulus, to the assertoric speech-act. And it is those utterances in the middle of this continuum that are

¹ There is talk of 'confessing' what one thought or dreamt on PI p. 222; which suggests a very limited class of past-tense avowals.

going to prove the most intractable. Wittgenstein asks whether the words 'I am afraid' are a description of a state of mind:

I say 'I am afraid'; someone else asks me: 'What was that? A cry of fear; or do you want to tell me how you feel; or is it a reflection on your present state?' - Could I always give him a clear answer? Could I never give him one? ...

... A cry is not a description. But there are transitions. And the words 'I am afraid' may approximate, more or less, to being a cry. They may come quite close to this and also be far removed from it (PI pp. 187 and 189).

Now 'a cry of fear' is clearly not an assertion; and if 'I am afraid' is uttered after reflection on one's present state, it may plausibly be said to constitute a self-ascription made on a behavioural basis. (Perhaps the object of fear is unconscious; at any rate I will have to think 'Look at how I have been behaving - I must be afraid of something'. Anxiety very commonly is self-ascribed in this way.) The problem-case is 'I am afraid' when it is properly described as 'telling someone how I feel'. The paradox of describing an utterance simultaneously as expressive and as intended to be informative is implicit in the following quotation, in which Wittgenstein then goes on to revel in the paradox!:

When it is said in a funeral oration 'We mourn our ...' this is surely supposed to be an expression of mourning; not to tell anything to those present. But in a prayer at the grave these words would in a way be used to tell someone something. But here is a problem: a cry which cannot be called a description, which is more primitive than any description, for all that serves as a description of the inner life (PI p. 189).

It is clearly time to set sail from the shores of Wittgensteinian paradox; but before we consult the very meagre navigational aids and weigh anchor, a few comments on the three kinds of case just discussed are in order. That category of avowals which 'approximate' to being cries, etc. - that are involuntary or at least more like reactions than intentional actions - will not be large. It would be more usual for the 22-carat expressive utterance

to take the form of an elliptical exclamation or ejaculation - 'Great!' instead of 'I'm really enjoying this!', etc. But the expressive theorist is really going to have to insist that when I tell someone how I feel, or report my mental state, my utterance is here also genuinely and purely expressive. The non-pernicious alternative to such a line would be to extend the class of assertoric behavioural self-ascriptions, in the following way. The basis, it may be said, of a large class of self-ascriptions beyond the narrow class of genuinely expressive avowals, is not anything Cartesianly introspected but simply one's disposition, if asked, to sincerely avow that very state. Well, there will be difficulties in characterising this disposition - would I have to specify conditions which prevented my sincerely responding to any questioning for instance? (Would it simply be a matter of being prevented here, rather than choosing?) But this dispositional approach in any case puts the cart before the horse; I don't say 'I am in pain' because I recognise I'm disposed to say it, but because I'm in pain! It is thus a rather desperate assertoric alternative.

The way to acceptance of the expressive thesis will be eased if one comes to recognise that controlled behaviour may be just as expressive as uncontrolled behaviour. The assumption that it isn't, is analogous to that viewpoint in aesthetics which sees truest self-expression as consisting in self-abandonment. (A viewpoint criticised in the present writer's Bill Evans - The Village Vanguard Sessions [1986].) Adherence to this viewpoint will result in failure to perceive much avant-garde artistic endeavour for what it is - the purest self-indulgence.

Finally, it should be noted that, at the other end of the spectrum, the range of behaviourally self-ascribed tokens of avowal-types is small. Some avowal-types, such as 'I am in pain', have no such tokens. In a few other cases, there is genuine ambiguity between avowal and non-avowal tokens - e.g. 'I am depressed' when it merely expresses a melancholic mood and when it aspires to the status of a clinical judgement (in the latter case, 'depressed' will have a technical sense).

We will return later to the notion of 'expression', because it is a key one in the resolution of the problem of avowals. What will

now be discussed is two kinds of criticism of a non-assertoric thesis. First, one directed at Wittgenstein's expressive account - the charge that it is in fact behaviourist. Second, the troubling question of the existence of truth-value links between avowals and their other-personal transformations. In the course of the latter discussion, the varieties of non-assertoric thesis, and the characteristics of the thesis which are doing the anti-Cartesian work, will come more clearly into focus.

4.7) Problems Faced by the Non-Assertoric Thesis(i) 'Covert Behaviourism'

What then of the charge of behaviourism? C. Grant Luckhardt's article 'Wittgenstein and Behaviourism' ([1983] pp. 318-38) is very clear on this matter. Luckhardt thinks that it is his account of avowals rather than of their 3rd-personal transformations which Wittgenstein fears will lead to the charge that he is a 'behaviourist in disguise' (*ibid.* p. 330). For if 'I am in pain' is merely expressive, and does not involve reference to a (private) object, then it may be thought that 'pain' itself has been 'cancelled out' and all that remains is the behaviour.

Well, this is a problem and a famous one, and it inspires some of Wittgenstein's most subtle persuasion. In fact, he does not see the problem as arising particularly from the analysis of avowals; rather from the general attack on the privacy of experience (see e.g. PI 293). But the expressive thesis of avowals is part of the resultant anti-Cartesian package, and it is worth examining the worry about behaviourism a little further. Luckhardt is right to emphasise that the conclusion of the 'beetle in the box' episode is a conditional:

... if we construe the grammar of the expression of sensation on the model of 'object and designation' [then] the object drops out of consideration as irrelevant (PI 293 - my emphasis).

It is this point that Wittgenstein repeatedly presses against the objection that he is a covert behaviourist; viz. that it is only on his account that pain is 'relevant' to the language we use, and that the behaviourist and the Cartesian have a common inability to make it so. How? Wittgenstein first insists that there could be no greater difference than that between pain-behaviour accompanied by pain and pain-behaviour without it:

... "And yet you again and again reach the conclusion that the

sensation itself is a nothing" - Not at all. It is not a something, but not a nothing either! The conclusion was only that a nothing would serve just as well as a something about which nothing could be said (PI 304).

To the behaviourist, pain is simply pain-behaviour; the sensation itself is 'a nothing'. But it will not be made to figure in a public language by being made to refer to a private object, as the Cartesian wishes; it is only by being logically connected to its expression that pain can be a topic of public discourse, and it is this connection which the Cartesian is in effect denying. The passage just quoted immediately precedes a comment which is often cited in connection with the expressive account of avowals:

The paradox [that pain is neither a 'something' nor a 'nothing'] disappears only if we make a radical break with the idea that language always functions in one way, always serves the same purpose: to convey thoughts - which may be about houses, pains, good and evil, or anything else you please (ibid. 304¹).

Wittgenstein later expands on the 'idea' with which we must make a radical break:

... you regard it as too much a matter of course that one can tell anything to anyone. That is to say: we are so much accustomed to communication through language, in conversation, that it looks to us as if the whole point of communication lay in this: someone else grasps the sense of my words - which is something mental: he as it were takes it into his own mind. If he then does something further with it as well, that is no part of the immediate purpose of language.

One would like to say "Telling brings it about that he knows that I am in pain; it produces this mental phenomenon; everything else is inessential to the telling". As for what this queer phenomenon of knowledge is - there is time enough for that ... (PI 363).

¹ cf. ibid. 317.

A familiar dialectic is in operation here - viz. that just because two possibilities are exclusive does not imply that they are exhaustive. 'Pain' in 'I am in pain' does not refer to a private object (a 'something'), but this does not mean it refers simply to pain-behaviour (leaving pain itself as a 'nothing'). There is another possibility undermining these two; that the utterance is expressive, that pain is expressed but not referred to by means of it. So it is not that the expressive account provokes a charge of behaviourism which needs to be countered, as Luckhardt maintains; but rather that the expressive account is part of a response to that charge. Having said this, the interpretation of the last two passages quoted is far from straightforward! Clearly, they both point to the hold which the descriptive model has on our thinking about language. Wittgenstein is concerned to point out the importance of other modes of speech-act; in the case of 'I am in pain', expression and response (assistance, sympathy etc.) rather than the mere transmission and receipt of information. On many occasions when I am in pain, I do not (perhaps cannot) tell someone else that I am; I just express it. And they just react - rather than, say, inferring that they ought to respond in certain ways. This, it seems to me, is at least part of the point of the cryptic para. 363.¹

If pain is not a private object, nor simply pain-behaviour, what is it? Insofar as this question does not mislead (see PI 308), one can say the following. Pain is not logically independent of behaviour, in two senses:

- (i) on particular occasions, certain verbal and non-verbal behaviour will constitute criteria for the ascription of pain.
- (ii) in general, one could not ascribe pain and other mental states to things which did not behave like human beings.

The latter view is a sort of 'truth in behaviourism' expressed by Wittgenstein in order to contrast with cruder manifestations of that

¹ See also pp. 225-26.

outlook:

"But doesn't what you say come to this: that there is no pain, for example, without pain-behaviour?" - It comes to this: only of a living human being and what resembles (behaves like) a living human being can one say: it has sensations; it sees; is blind; is deaf; is conscious or unconscious (PI 281).

Also:

But do I also say in my own case that I am saying something to myself, because I am behaving in such-and-such a way? I do not say it from observation of my behaviour. But it only makes sense because I do behave in this way (ibid. 357).

As Luckhardt indicates, it is Wittgenstein's view that reports of mental states are made through reports of behaviour:

"I noticed he was out of humour". Is this a report about his behaviour or about his state of mind? ("The sky looks threatening": is this about the present, or the future?) Both; not side-by-side, however, but about the one via the other (ibid. p. 179).¹

But to comment any further on this highly suggestive passage would be to anticipate a projected work on the Arbitrariness of the Mental/Physical Distinction.

(ii) Truth-value links (I): How could avowals subserve inferences?

The most serious problem faced by the non-assertoric thesis, however, is the following: how can a sentence used to effect a non-assertoric utterance liaise with assertions (with which, one would like to say, it has 'truth-value links'), and in particular

¹ cf. ibid. p. 223: "'What is internal is hidden from us" - The future is hidden from us. But does the astronomer think like this when he calculates an eclipse of the sun?'

subserve inferences? The 'simple-minded' picture of the role of tense, whereby the making of the same statement at different times is facilitated, generates an objection that might be dismissed with a Wittgensteinian snort:

"But surely 'I believed' must tell of the same thing in the past as 'I believe' in the present!" - Surely $\sqrt{-1}$ must mean just the same in relation to -1 , as $\sqrt{1}$ means in relation to 1 ! This means nothing at all (PI p. 190).

(That is, roughly: the ' $\sqrt{\quad}$ ' sign converts 1 into a smaller rational number - but it doesn't do the same to -1 , nothing like. 'I' converts '... believed' into a self-description when completed with a 'that' - clause. But it doesn't do the same to '... believe'.) Anscombe gave a similar snort when interrogated by the present writer over truth-value links between 1st- and 3rd-person statements (as related in Elizabeth Anscombe und der Bangor Kreis; see p. 237). She in effect pointed out that co-reference is the ('natural') explanation not the datum. Now the same is true with respect to the relation of avowals to their 3rd-personal and differently-tensed transforms. The datum here may be different though - not inter-substitution salva veritate, since that description presupposes the falsity of the non-assertoric thesis, but rather either some full and tedious description of the phenomenon, or maybe inter-substitution 'preserving appropriateness of sincere utterance'. The kind of riposte expressed here - that 'compositionality alone doesn't take care of content' - is a familiar Wittgensteinian one, occurring also for instance in the discussion of 'It's 5 o'clock on the sun' (PI 350). But is it good enough in the present context?

The more particular objection relating to inference is discussed in Blackburn ([1984] pp. 189-196), in the context of a defence of quasi-realism. (It originates with an article by Geach [1964], which utilises a point by Frege.) The argument Blackburn states is this, relevantly amended.¹ 'I am in pain' may occur unasserted, in contexts making up larger sentences. Yet it must

¹ Blackburn is concerned with an expressive account of moral judgements rather than of avowals.

mean the same both embedded and on its own. This is because 'I am in pain' must subserve inferences, and if there were no identity of meaning a fallacy of equivocation would be committed. Consider the following:

I am in pain

If I am in pain I deserve sympathy

So: I deserve sympathy.

Anyone asserting the second, hypothetical premise is not expressing pain. Yet, so it is said, the meaning must be the same for the occurrence of 'I am in pain' as the first premise and as embedded in the second. This cannot be secured, and fallaciousness avoided, if 'I am in pain' is, in non-embedded occurrences, purely expressive.

Well, there is a problem for the non-assertoric theorist in the vicinity, but this is not it - as is in fact apparent from Blackburn's subsequent discussion. The problem of equivocation is a quite general one which will arise for any inference with a conditional premise, unless one separates force and propositional content. Thus, the equivocation would arise too for an assertoric account of 'I am in pain', since anyone asserting the second, hypothetical premise is not asserting 'I am in pain' either.

In fact it was this problem amongst others which prompted Frege to introduce the distinction between force and content. Let's start with the assertoric account of 'I am in pain'; the argument may then be represented thus (where ' \vdash ' is the sign for assertoric force):

$\vdash(P)$

$\vdash(P \rightarrow Q)$

$\vdash(Q)$

Since the ' \vdash ' sign does not represent any part of the meaning of either ' $\vdash(P)$ ' or ' $\vdash(P \rightarrow Q)$ ', there is no equivocation in meaning between the two occurrences of 'P'.

Likewise there will be no equivocation if there is a similar separation in the expressive account (where '!' is the sign for

'expressive force'):

!(P)

!(P → Q)

!(Q)

There is no problem of equivocation unique to the expressive account, then, it seems - but other difficulties do arise! One might, for instance, wonder how a non-assertion could figure as the premise of an argument. Or (relatedly) there is the Geach point that Blackburn goes on to discuss in connection with the quasi-realist account of moral judgement - viz. it seems that the antecedent of a conditional needs to be able to stand as an assertion. For how could a mere expression of pain have an implication?

(iii) Truth-value links (II): The Haldane Objection

Fortunately, there is no need here to go into the issue of the conditional form which Blackburn discusses at some length. For even if the question were resolved in favour of the expressive account, there is a more serious objection, which I term the Haldane Objection, applicable specifically to an expressive account of avowals which makes use of the force/content distinction to surmount the equivocation problem.¹ This is as follows: 'By introducing an expressive force operator '!', and applying it to avowals in place of an assertoric force operator, the central problem has just been pushed one step back. In seeking to defuse the truth-value links problem by asserting a common propositional content between different occurrences of "I am in pain", the Trojan still needs to give an account of that content, with its apparently inescapably Cartesian truth-condition'.

The Haldane Objection opens up a dilemma which shows that after all there is a problem of equivocation unique to expressive accounts. For the non-assertoric theorist must either assert a common propositional content between expressive and non-expressive

¹ It was suggested to me by John Haldane.

uses (in which case he faces the Objection) or else he must go for a more radical option in which the original problem will re-appear. The proponent of this latter option will say of the former:

It assumes that corresponding to the characteristic expression of 'I am in pain' there is a thought to be had - that I am in pain. (It is just that it cannot be asserted.) But in fact it is an implication of the non-assertoric thesis that corresponding to the truly self-conscious expression of 'I am in pain' (rather than any possible behavioural self-ascription) there is no thought to be had. Consequently 'I am in pain', though expressive, does not have 'expressive force' - nor any other kind of force.

It is indeed an attractive idea to transcend the Haldane Objection in this way (though not, presumably, to those wedded to the idea of a systematic theory of meaning). But doesn't it open up again the problem stated at the outset that Blackburn raises - the problem of equivocation - and in just the same terms?

Well, it is open for the radical Trojan to reply as follows: The use of 'I am in pain' in the first premise is clearly not expressive. But there is no behavioural self-ascription here, so how can it be assertoric either? Well, the argument is not one I ever seriously make use of; but if I am called upon to spell it out, I can represent the first premise as an assertion, verifiable in a way in which it never seriously is verified - behaviourally. I don't ever seriously go through the steps outlined in the argument but the charade of doing so could involve the charade of observing my own behaviour to check whether I am indeed in pain. And the argument could be represented in the Fregean manner (p. 200) to avoid equivocation. Thus the sense of unreality and triviality about the objection is maintained in the response to it.¹ (The Geach

¹ P.M.S. Hacker seems to have a similar point in mind in his revised thoughts on truth-value links objections. He writes that 'these ... objections are exemplary cases of being so mesmerised by forms of words that one becomes oblivious to their use ... one forgets to examine the point of employing

point about the conditional is catered for too, incidentally.)

This is not of course to say that the argument with 'I am in pain' is at all defective as an argument. There are innumerable trivial arguments where the epistemic possibility of someone recognising the truth of the premises but not that of the conclusion is not open. One might contend that in such cases, since the argument cannot be intended as a piece of persuasion, it is hard to say what use it has. Why then should one insist that the constituent premises should have their normal use or force? Why not fix on some arbitrary use that satisfies the requirements of logic? ('Avowal' is a term of pragmatics, applying to sentence-tokens and not types.)

The picture of avowals that provoked this response - that they are structured but without force - will now be explored in some detail. But the reader should bear in mind that if the above response fails to satisfy, all is not lost for the expressive account - since it will subsequently be questioned whether an expressive account need be non-assertoric at all.

the resultant sentences' ([1986] p. 303 - see Hamilton [1987 - forthcoming]).

4.8) Do we need a Non-Assertoric Thesis anyway?(i) 'Non-force theses'

The presentation of the Haldane Objection (p. 201) has put us in a position to distinguish the three essential varieties of non-assertoric thesis. Firstly, as has already been mentioned, there is the distinction between theses of structured and unstructured utterance. It is now clear, however, that there is a further distinction between two kinds of structure - that proposing a common propositional content between avowals and non-avowals, and that denying it. The Haldane Objection leads us to view the second option more favourably than the first. It looks as if we do not want to view 'expression' as a kind of force attaching to a propositional content; but can we deny that this is what it is, without thereby denying the possibility of structure altogether? Two possibilities present themselves at this point:

(i) avowals are non-assertoric and purely expressive; they possess a structure but no force component.

(ii) avowals are, whilst expressive, also assertoric; but they are however non-cognitive.

The second account, assertoric but non-cognitive, constitutes an attempt to resolve the problem of authority by extracting what is of value in the non-assertoric thesis (including its emphasis on expression) whilst by-passing its difficulties. In the case of non-assertoric variant (i), even, these difficulties are formidable. Whilst some of the difficulties will apply to any version of (i), we should distinguish two versions at the outset:

(i)' avowals do not possess force; nor does anything else.

(i)'' avowals constitute an enclave of force-less speech-acts in a general field of force.

These are really more properly called 'non-force' rather than non-assertoric theses - though (i)', the radical non-force thesis, has as its effective target the notion of assertion, installed in its place as the paradigm type of force. It is possible that the latter thesis is that espoused by Wittgenstein in the Investigations - paras. 22-24, and also in those paragraphs we have already discussed on p. 196.¹

The radical non-force thesis takes us beyond the question of avowals into a central question in the philosophy of language which it is not possible to discuss in the present context, except to say the following. The thesis maintains that there is no unified notion of 'assertion' - it cannot be characterised in terms of necessary and sufficient conditions nor in any other terms that would make it a profitable notion to employ. Rather, the class of 'assertions' is fragmentary with no interesting features in common. Now, this thesis does not have as an implication that the authority of avowals is no longer a puzzling phenomenon. It needs to be made clear that the advantage of a non-assertoric thesis is that it rules out (or at least makes ludicrously inapposite) a Cartesian explanation of that authority, without doing away with the need for an alternative explanation. The pertinent point here is this: it is the fragmentary notion of authority, which goes with the fragmentary notion of assertion, that is of central importance. And this, it seems, is available without recourse to the radical non-force thesis (and arguably, but this is to anticipate, without recourse even to a non-assertoric thesis at all).

Less upset to received ideas is caused by the less radical thesis (i)', and it is to this that we now turn, in doing so expanding on the points just made. The 'fragmentary' notion of authority allows for, in fact requires, the idea of an authority not based on expertise, being in the best position to judge, assessment

¹ It is true that, in criticising Frege, Wittgenstein's remarks are directed as much at the former's residual psychologism about meaning as at the notion of force (assertoric force in particular) per se. Dummett discusses the matter sensibly in his [1981] pp. 359-62.

of evidence, etc. But such an idea, if it can be made out, is also available for an enclave of force-less speech-acts (avowals and others) in a general field of force. That is, the very same idea of authority required by the radical thesis (i)', is also available for avowals, under thesis (i)'' . And this, if demonstrable, will satisfy the key requirement for any satisfactory account of avowals - that their authority does not involve expert knowledge. (Their authority will not only need to be described, however, but also explained.) Discussion of the 'enclave' thesis (i)'' will be facilitated by an analogy pleasing to cricket-lovers (though imperfect). When an umpire in cricket gives a batsman out, or (better, since the former kind of decision is signalled non-verbally) cautions a bowler for running on to the pitch in his follow-through, his speech-act is not available to be made by someone else. The judgement is normative or constitutive within the game, though criticisable by others at least insofar as they may question whether it was warranted by the physical basis on which it was made. (Not criticisable in the sense that players may reject the decision, of course.) It may be reported by others, who may also report its physical basis (the bowler persistently ran onto such-and-such an area of the wicket). But they cannot actually say, 'I'm warning you for running on to the pitch', and achieve the same kind of speech-act as the umpire achieves (i.e. delivering a warning). This is not because the utterance possesses a propositional content inaccessible to anyone else, but because the speaker, when he is the umpire, has a certain institutional role, such that the utterance of those words by him creates a particular institutional fact.

J.L. Austin, in his [1962], made a distinction between performatives and constatives. He then discovered that the constatives were performatives too - that one did something by making a constative utterance also. The present suggestion is that whilst 'constatives' are performative, some utterances are purely performative. Examples would be: 'I'm warning you for running on to the pitch', 'I thee wed', 'I name this ship the "Lollipop"' - and 'I am in pain'. These utterances do not exhibit the normal components of propositional content and force. They effect speech-acts not available to be effected by anyone else in the situation in question; not because they possess some propositional content truly

accessible only to the speaker,¹ but because only the speaker has the institutional role that enables him to make the speech-act and thereby determine (or help to determine) the institutional fact in question (that the player was cautioned, that the ship was named 'Lollipop', that the couple were married - and that my pain was avowed). On the view presently being considered, then, 'only I have the authority to avow my mental states' (with all that that entails in terms of the Famous Five) is comparable to 'only the umpire has the authority to caution the players'.

This description of avowals as 'pure performatives' analogous to e.g. umpiring decisions, opens up the possibility that their authority may be explained. What we would be looking for would be an explanation of the advantages of the practice, along the lines of the 'sociological' explanation of the advantages of having an umpire in a competitive sport. However, two problems with our analogy surface before we may consider such an explanation.

Firstly, it will have to be insisted, by the proponent of this thesis (i) whether radical or not, that 'no force' does not mean 'no structure'. And this indeed is plausible. In rejecting force for 'performatives' we are indeed posing problems regarding their integration into a systematic theory of meaning; but as far as less parochial requirements of structure are concerned, there seems little problem. The Generality Constraint may be satisfied given that an understanding of the individual components of the performative may be exhibited in other kinds of sentence, purely performative and otherwise. The embedding-objection may be parried along the lines already suggested, by means of a Wittgensteinian snort and by means of the 'charade' response (pp. 199-203).

Secondly, the analogy between an umpire's decision, and my avowals as examples of expressions of authority lacking in linguistic force, is imperfect of course. One reason is that avowals, unlike umpiring decisions, can be 'defeated' (we may discover the subject was insincere, to take the principal instance). (An analogy with tennis where the tournament referee (Capt. Mike

¹ The Cartesian view, for our purposes.

Gibson) can over-rule the umpire, was considered by the present writer but rejected because of the appalling commercialisation of the game and its pollution by chiefly American vulgarity.) A second and more important reason, is that the 'institutional setting' is much more explicit in the case of sporting compared with avowing practices. Finally, and possibly fatally, umpiring judgements are not groundless, but based on observable facts which the umpire is in the best position to note. In this crucial respect then, the kind of authority expressed differs.

(ii) A non-cognitive thesis on its own

The problems of embedding (if the Wittgenstein snort and associated responses fail to impress) and the dubious extent to which non-assertoric status per se (rather than any non-expert notion of authority attached to it) is actually required for the resolution of the problem of avowals, may encourage us to distance ourselves yet further from Trojan orthodoxy, and embrace the non-cognitive thesis of avowals alone. (This is, insofar as these crude labels mean anything any more, a non-Trojan position.) It should be clear (though it has gone so far undiscussed) that the non-cognitive thesis of avowals is a consequence of the non-assertoric thesis. What we are now asking is 'can they be separated?' Can the non-cognitive thesis be maintained compatibly with at least an agnosticism about the assertoric status of avowals? Wittgenstein's own support for the non-assertoric thesis has been questioned; and though many of the doubts cast on his sympathy for that line of approach can be dispelled by careful formulation of the different varieties of non-assertoric thesis, commentators have been right to view his commitment to the non-cognitive thesis as less debatable. There is less doubt in the case of the latter partly because the thesis is easier to understand.

The endorsement of the non-cognitive thesis in Wittgenstein's writings is clear. One might expect nuances analogous to those apparent in his expression of what seemed to be a non-assertoric thesis; viz. that there is a range of cases, that on occasion it might make sense for me to say that I know I am in pain (when it is correct to say that I am describing rather than expressing my pain).

But in fact 'I know that I am in pain' is ruled out as having no use at all.¹ This view goes back at least as far as the Blue Book (and, pace Hacker, the evidence of its being denied in the Philosophical Remarks is quite unclear²). There, it becomes clear why it is easier to ascribe the non-cognitive thesis to Wittgenstein; his presentation is unequivocal, and employs a familiar distinction between grammatical and empirical propositions (or here, 'metaphysical' and 'experimental'):

... when in a metaphysical sense I say 'I must always know when I have pain', this simply makes the word 'know' redundant; and instead of 'I know that I have pain', I can simply say 'I have pain' (ibid. p. 55).

In the Investigations, the non-cognitive thesis is, in a famous passage, explicitly linked with the relocation of the Other Minds problem as the 'Non-Other Minds Problems' (to use R.S. Buck's inelegant but handy phrase). Wittgenstein denies that sensations are in any sense 'private':

If we are using the word 'to know' as it is normally used (and how else are we to use it?) then other people very often know when I am in pain - Yes, but all the same not with the certainty with which I know it myself! - It can't be said of me at all (except perhaps as a joke) that I know I am in pain. What is it supposed to mean - except perhaps that I am in pain? (ibid. para. 246).

As Geoff Phillips has implied in discussion, commentators have quite failed to explain what the 'joke' could be here, and indeed we must set aside the curious parenthetical qualification as an example of the author's peculiar sense of humour, and simply say that 'I know I am in pain' has no use, on the Wittgensteinian view.³ The

¹ Indeed, Wittgenstein writes that 'even when someone says ('I am frightened') as a piece of information, he does not learn it from his sensations' (PI p. 174).

² See his [1972], p. 254, and PR 66.

³ This latter point will not however carry weight until it

assault on Cartesianism here is proceeding on two fronts - or rather, the assault on the cognitive account of avowals is facilitating the dissolution of the Other Minds problem, which otherwise would be intractable. There is no superior 1st-personal knowledge the standards of which 3rd-personal knowledge fails, to its detriment, to reach:

Other people cannot be said to learn of my sensations only from my behaviour - for I cannot be said to learn of them. I have them (ibid. para. 246).

The Blue Book distinction is implicit in the next paragraph:

'Only you can know if you had that intention'. One might tell someone this when one was explaining the meaning of the word 'intention' to him. For then it means: that is how we use it. (And here 'know' means that the expression of uncertainty is senseless.) (ibid. p. 224)

'Only you can know' is a 'grammatical proposition', then; such expressions have a constitutive role in the 'language-game', and should be treated accordingly:

'He alone knows his motives' - that is an expression of the fact that we ask him what his motives are (ibid. p. 224).

('Desires' would have been a better example than 'motives' here, since whether the subject has authority on the latter is more questionable.)

The non-cognitive thesis is, in the Investigations, not only linked expressly with the resolution or relocation of the 'Other Minds' problem, but is also clearly connected with that account of knowledge which receives its fullest expression in On Certainty. In the Investigations, after dismissing as 'philosopher's nonsense' 'I know what I want, wish, believe, feel ...', Wittgenstein writes:

is shown that it is a genuinely semantic and not purely pragmatic feature.

'I know ...' may mean 'I do not doubt ...' but does not mean that the words 'I doubt ...' are senseless, that doubt is logically excluded (ibid. p. 221).

In contrast to the Cartesian, and to the Moore criticised in On Certainty, who each aspires to the foundation of empirical knowledge in self-evident truths, Wittgenstein maintains that knowledge implies at least the logical possibility of doubt and is inseparably connected to the notion of evidence - not to self-evidence (e.g. OC 504). It is a familiar dialectic in which the non-cognitive thesis is embedded - the famous '"exclusive" doesn't imply "exhaustive"'. Just because I cannot doubt whether I am in pain, does not mean I must know that I am in pain - rather, there is no cognitive status here at all. We may round off this homely exposition with a famous and telling metaphor:

It is correct to say 'I know what you are thinking' and wrong to say 'I know what I am thinking'.

(A whole cloud of philosophy condensed into a drop of grammar)
(PI p. 222).

(iii) Even a non-cognitive thesis is unnecessary

Let us assume the non-cognitive thesis can be held in isolation from its non-assertoric counterpart. What would be the advantage of so doing? Well, the thesis is less extreme, and yet delivers all the advantages of the non-assertoric thesis without many of the bothersome complications of the latter. These advantages can be summed up thus: the theses encourage a non-Cartesian account of the authority of avowals. The question now arises: is the non-cognitive thesis itself actually necessary for such an account? We can adumbrate two accounts which suggest not. First, it may be argued that the subject may be said to know that e.g. he is in pain, provided that the avowal is not seen as an expression of that knowledge, as the product of some cognitive process. The question of the subject's knowledge is not one that arises - it is not asserted as a denial of someone else's claim that he doesn't know, for instance. His knowledge may, however, be expressed by means of some such peculiar utterance as 'I know I am

in pain'. How someone may come to make such an utterance may be quite opaque. But the important point is that the authoritative utterance 'I am in pain' is not an expression of my knowledge that I am in pain, in the way that my utterance of e.g. 'The Famous Five characterise avowals' is an expression of my knowledge of the essential rightness of our general approach, say. A better account says that there are two sorts of knowledge - that which involves some sort of cognitive achievement and that which doesn't (in some way yet to be described). We may thus say what we like (all the things which it seems natural to say) about the subject's knowledge - provided we appreciate that the kind of knowledge he has is of the sort involving no cognitive achievement. This better account we will investigate shortly.

We have considered proposals (i) and (ii) of p. 204: the non-force thesis and the non-cognitive thesis. All that is left of these proposals now is the idea that the authority of avowals is somehow analogous to that of certain kinds of institutionally-conferred authority. We have already on pp. 207-8 had reason to question the appropriateness of this kind of analogy; and the business of providing an explanation of the authority of avowals, along the lines of the 'sociological' explanations in the other cases, has yet to be embarked on. Without such an explanation, of course, the view canvassed would stand condemned as the conventionalist view of p. 147. The difficulty of providing any such explanation, however, is a symptom of the fact that the analogy itself is desperately unconvincing - for reasons already given. The disanalogies there indicated are reinforced by a gut reaction that this explanation of the authority of avowals could not be 'deep' enough; is my authority really only comparable to that of some official? This reaction may have a Cartesian provenance; but it need not.

The state of play at present is as follows, then. First, doubt has been cast on the necessity both of a non-assertoric, and of a more limited, non-cognitive account, for any successful resolution of the problem of avowals. Second, the authority of avowals does not seem to be profitably compared with the different kinds of institutionally-conferred authority. Thus an impasse has been arrived at. The tentative or negative tone of the preceding discussion will now be atoned for by the attempt to unify the

accounts of the cognitive and the authoritative status of avowals around the concept of a subject of experience. Our ship, having set sail from the shores of Wittgensteinian paradox on p. 192 with meagre navigational aids and since then having floundered in choppy seas before becoming becalmed, should now be set on course - with the account of expression being rendered less paradoxical.

4.9) Resolution (I): Knowledge and Authority(i) Subject and expression

The 'unified account' is suggested by the following considerations. We need to explain why the sincere avowal guarantees the truth of the corresponding 3rd-person ascription. 'Why should the 1st-person utterance be accorded a special weight denied to the 3rd-person utterance?' it is asked. Well, to ask this is to assume one is (in an important sense) comparing like with like. But it is not as if one is comparing 'one person's view' with 'another's'; that would be to confuse a description of a phenomenon ('He is in pain') with something that is part of the phenomenon itself (the avowal). To take the Wittgensteinian idea of 'expression' seriously is to see that the avowal is logically connected with the pain, or the mood, or whatever. If there were no expression, there would be no sub-
jects of experience; all would be mere observers (whether of their own 'mental states' (so-called) or of those of others). 'Experience' goes with 'expression' as 'observation' goes with 'description'. As Evans writes in connection with perceptual experience:

[The subject's] internal state cannot in any sense become an object to him (He is in it) (ibid. p. 227).

In comparing 'I am in pain' with 'He is in pain' and asking why one is accorded special weight and the other not, one is, as it were, attempting to contrast two pieces of observation and asking why one should be privileged, and forgetting about a different sort of relation to that of observer and object, viz. that of the subject to the state he is in.¹ The vertiginous effects of thinking about

¹ The absurdity of treating one's own avowable states as objects of observation is brought out by Wittgenstein in the Investigations:

If you observe your own grief, which senses do you use to observe it? A particular sense; one that feels grief?

the latter relation should not lead us to reduce it to a special case of the former; rather, it is sui generis. Along with this relation goes the idea of a subject giving expression to the state he is in, in contrast merely to acquiring knowledge of it. Before dealing with 'expression', however, we must first say something about the subject's knowledge.

(ii) The 'thin' sense of 'knowledge'

When we put the notion of a subject of experience at the centre of our account in the way just adumbrated there will be two options available regarding the cognitive status of avowals - which, in our new 'unified' account, we are now treating in tandem with their authoritative status. The first option is to deny that the notion of knowledge (and consequently that of authority) has application with regard to my own avowable states - this amounts, of course, to the non-cognitive thesis. The alternative option is that suggested by the second, superior account of p. 212 - to distinguish between different kinds of knowledge (and now, authority) and maintain that the kind of knowledge I have of my own avowable states is of that kind involving no cognitive achievement. Before elaborating on these options, something needs to be said about the connection with authority. The connection is simply this: 'authority' is a cognitive notion, and it makes no sense to separate the grounds for its application from the grounds for the application of 'knowledge'.¹ Consequently, our already familiar first option, the non-cognitive thesis, should be supplemented with a 'non-authoritative thesis' of avowals. It makes no sense to talk of avowals as 'authoritative' when they are viewed as non-cognitive.

Then do you feel it differently when you are observing it? And what is the grief that you are observing - is it one which is there only while it is being observed? 'Observing' does not produce what is observed (ibid. p. 187).

¹ We are talking of authority in the sense of 'authority on' (some subject) rather than political authority, of course; which is not to deny important connections between these two senses.

We should rather stick simply at the level of neutral description (the 'Famous Five') in the manner in which the 'unified account' began - when it was said that what needed explanation was why the sincere avowal guaranteed the truth of the corresponding 3rd-person ascription. In contrast the second option, which is now favoured, in fact charts a 'middle way' which avoids the polarisation between Cartesianism and the residual Trojanism of the non-cognitive (and now, non-authoritative) account. As in many such middle ways, the desired effect is achieved by invoking a 'thin' sense of some key-concept or concepts - in this case, 'knowledge' and 'authority' (cf. the 'thin' sense of 'truth' in the realism - anti-realism debate). This sense will, in all probability, be applicable uniquely to one's knowledge of, and authority regarding, one's own avowable states. It is the knowledge the subject has, not of objects, but of states of himself. As was insisted above, the relation between the subject and the state he is in is sui generis and not a special case of that between observer and object; what is presently being canvassed is the idea that the knowledge the subject has of his avowable states is sui generis and not a special case of knowledge he has of objects. The triteness of the preceding remarks may be alleviated, in two ways. First, we must understand something of what the central aspects of the concept of knowledge in general are, in order that we may see what the 'thin' sense lacks. Second, we must see how the facts about self-knowledge mesh into the account of a 'thin' sense of 'knowledge'.

It is important to make clear how the preceding two tasks are related to the central question of explaining the Famous Five (to put the matter neutrally). They are related as follows: The claim that avowable states are knowable in a 'thin' sense by the subject has the same dialectical role as the claim that avowals are non-assertoric, or that they are non-cognitive. These claims rule out the wrong kind of explanation of the Famous Five (the Cartesian kind). None of them allows for the 'natural' kind of explanation in terms of cognitive achievement, privileged access, etc. But they do not thereby constitute an alternative explanation. The latter is something additional, and the only alternative explanation thus far considered is the attempt to draw together the authority of avowals and various kinds of institutionally-conferred authority. This did not work. Now, in conjunction with the account of avowable states

as knowable in a 'thin' sense, is proposed the explanation of the Famous Five in terms of the idea of a subject of experience giving expression to those states. This explanation is 'conceptual' in contrast to the 'pragmatic' explanation of the institutional account. The new 'cognitive / expressive' account is one which fits in better with our intuitions on the matter - and terminates that process whereby Trojanism has been stripped of all features thought to be required for the avoidance of Cartesian illusion but not in fact so required. (A process which began in Chapter 2 with the doubt cast on the 'no-reference' view of 'I'.) The end-product is one not essentially Trojan at all, of course; though as with many of the deepest controversies in philosophy, perhaps, the original labels increasingly seem to lack profitable application as the dispute becomes further removed from what is left of any well-grounded division of opinion.

How, then, is this 'thin' sense of knowledge, applicable to cases of knowledge of one's own avowable states, to be characterised? We need to note the existence of two central aspects of the concept of knowledge. First, there is the aspect of 'pedigree' - which enters essentially into any analysis of what it is for X to know that p. This of course is the aspect on which there is a heavy concentration in the literature, in which causal and other accounts vie with justificatory ones, each seeming to lack that 'vital ingredient'¹ necessary to provide a water-tight characterization. Second, there is the aspect of 'performance' - which enters importantly into any account of what is involved in avowing 'I know that p'. This aspect is an important part of the pragmatics of knowledge-claims over and above the purely semantic analysis of conditions for 3rd-personal ascription of knowledge. For in saying 'I know that p', an essential part of what I am doing is 'giving a guarantee'. I am assuring my listener of my confidence in what I am saying. This is the 'performative aspect' of knowledge-claims, the kind of force attached to, or comprising, the act of avowal.

In the case of self-knowledge regarding avowable states, there is only the aspect of 'performance'. There is no need for an aspect

¹ See Nozick [1981] pp. 169-71.

of 'pedigree'; there are no third and ensuing conditions to be secured before we have a case of genuine knowledge. That is - if I know that X is joyful, and I know that X believes that he is joyful, then I know that X knows that he is joyful. The 'third condition' (that X is justified in believing ..., that X is appropriately causally linked to the fact that ..., or whatever) has no place here. Now, it clearly makes no sense for us to suppose that X may offer a justification for his belief that he is joyful; but we may suppose an appropriate linkage of the sort advocated by causal theorists. However, such a causal linkage will not be a condition for distinguishing between cases of belief and cases of knowledge, since there is no such distinction to be made here.

So, where the sense of 'knowledge' is 'thin', only the two conditions (p and X believes that p) suffice for its application. Now, we have performed a little of our second task of alleviating the triteness of p. 216; showing how the facts about self-knowledge mesh into the account of a 'thin' sense of 'knowledge'. What needs to be added is simply that what has been said is another way of saying that self-knowledge of avowable states is 'immediate', 'non-inferential', 'non-observational'.

It is in the inter-connection of a 'thin' cognitive thesis of avowals and an account of their role as expressions of a subject's experience that the way forward lies to a resolution of our central problem. The aim of the latter account is simply to explain the Famous Five in a non-Cartesian way - which is, it has been argued, thus to explain the authority of avowals, in a 'thin' sense of 'authority'. But it should be noted that it is arguably a matter of one's predilections whether one believes it is authority that is being explained (and whether one believes avowals are cognitive). The role of the 'thin' cognitive thesis is partly to show that a non-Cartesian explanation of the Famous Five in terms of 'expression' need not be counter-intuitive with respect to the cognitive status of avowals - that it need not go against our feeling that I can and do know whether I am in pain, etc. But it is arguable that an 'expressive' explanation of the Famous Five, of the sort about to be outlined, is compatible with a non-cognitive thesis. Be that as it may, it has been argued that the non-cognitive thesis is not merely counter-intuitive but also unnecessary; an over-reaction

against a certain model of self-knowledge in terms of 'privileged access', 'best possible vantage-point' - it is the latter model which is objectionable. The 'thin' cognitive account is of course quite unacceptable to the Cartesian because (to him) insufficient.

(iii) Arguing against an 'observational' account of self-knowledge

Let us expand on the explanation adumbrated on p. 214. Avowable states, it was urged, are not objects to a subject of experience; he experiences them, but does not observe them. Avowals, consequently, are not reports of observation but expressions of experience.¹ Some perhaps may be satisfied with the plainness of this dichotomy, but surely it is in reality hard enough to understand, never mind to attempt to argue for.² Enlightenment can proceed from two policies. First, by attempting to argue against an 'observational' account of self-knowledge. Second, and more importantly, by attempting to outline the notion of 'expression' in such a way that the naturalness of its application to avowable states (in contrast to that of observation) is perceived. The arguments involved in the first policy seem to display varying degrees of crudeness - partly because we are still dealing with as yet unrefined notions. An initial argument could be termed that of 'unacceptable attenuation of the subject'. If I can observe my own avowable states, what is the status of this observing subject? What properties does it possess? Presumably not the properties of being in pain, being sad, etc., that it is observing. The subject, then, would be attenuated to a propertyless point, a mere insubstantial perspective. Two complaints might be made here. First, even if we accept the argument's presumption, does it furnish a reductio ad absurdum? And secondly, is the presumption correct - must

¹ Putative avowals which tend towards the descriptive will simultaneously tend to depart from the exhibition of the Famous Five.

² The dichotomy is hardly novel; it is that, for instance, in Ryle's Concept of Mind, which goes under the heading 'Sensation and Observation' (pp. 190-231). But it is there inadequately developed and argued for.

observation of pain by a subject exclude self-predication of pain by that subject? I won't dwell on the first point - it is one congenial to Trojanism, which is after all very much interested in the attenuation of the subject. The second point indicates that really the 'unacceptable attenuation' argument is a corollary depending for its effectiveness on a prior argument - that experiencing (being in) an avowable state is incompatible with observing that state. The kind of argument we're looking for here would be analogous to that offered by Von Wright in his [1971] to the effect that agency excludes observation of the causes of one's own basic actions. This, however, is a line of enquiry on which I will say no more - since the second of the policies above seems more promising.

4.10) Resolution (II): 'Expression' (II)(i) Taylor's account

Let us then proceed on the second task, by examining, in positive fashion, the contrasting notion to that of observation - 'expression'. We can begin by stepping back a little and locating this notion in the wider context of a philosophy of mind of Wittgensteinian cast. It has been observed at several points that Wittgenstein re-located the Other Minds problem in a 'Problem of Non-Other Minds'. We can now improve on this statement by saying that adopting a Cartesian stance will inevitably result in intractable difficulties in one area or the other; and that the development of certain superficially disparate ideas in the Investigations may offer a common solution to each sort of problem. The connection may be put laconically as follows: 'seeing - as' is not inferring just as expressing is not reporting. The essentially public nature of avowable states alike rules out an inferential model of our knowledge of other minds and an observational model of our knowledge of our own minds. Of course, just as there are some kinds of sceptic whom the notion of a 'criterion' fails to satisfy, so it would be foolish to claim that the notion of 'expression' will achieve general approval; but by attempting to demonstrate its centrality, one will at least be showing the essential unity of the problems of knowledge of self and of others, and of the only adequate sorts of response to them.

How might this result be achieved? Use will need to be made of ideas in Charles Taylor's estimable [1979] - an article which contributes to his salutary efforts to demonstrate the impoverishing effects of the causal paradigm on theories in the philosophy of mind. On pp. 73-78, Taylor offers various conditions constraining 'expression' in the sense in which action could be an expression of desire. But his account seems to have a more general application at least to the range of paradigm-avowals as expressions of avowable

states in the required strong sense.¹ It runs as follows:

- (i) the kind of manifestation which is an expression must at least offer a 'physiognomic reading'; it must make something manifest 'in an embodiment' - that is, 'directly, not leaning on an inference' (ibid. pp. 73-4).

Thus X may be manifest in Y, where X is not identical to Y, and yet Y not be an expression of X. For instance, the chalk dust on your lapels makes manifest the fact that you've been near a blackboard, probably writing on it, perhaps that you're a teacher. But it doesn't express any of these things; since I infer to them from the chalk dust. In contrast, I do not infer from your facial expression to the joy which it expresses.² If your hand trembles I may infer, on the basis of my experience of you, that you are agitated - this will be a symptom as opposed to the criterion in the joy example; of this and other tricky issues in the Taylor account, see below.

However, some cases of 'physiognomic reading' in the above sense are not cases of expression in the required (strong) sense, nor maybe in any sense. Taylor gives the example of reading its impending fall in an upright construction. I can characterise my grounds for the former only in terms logically connected with it, e.g. 'instability'; here, then, is a genuine 'physiognomic reading', yet we would not want to say that the unstable construction expresses its impending fall. Two further conditions are required, then, the first fairly straightforward:

- (ii) In mere physiognomic reading, the X we read in Y can be observed on its own. In cases of genuine expression, what is expressed can be manifest only in that or some other expression.

¹ Taylor mentions the possibility of viewing linguistic behaviour as expressive in a strong sense, but does not elaborate on this and in particular does not discuss avowals.

² Some may demur. But the very great difficulties of arguing for an inferential account here should be well-known.

I can do more than read the impending fall in the appearance of the construction; I can, aside from a physiognomic reading, see it when it actually collapses. However, the joy you see in my face, or to take a different example, the sadness the music expresses, could be manifested alternatively only in some other expression - I might dance for joy, that particular mood might also be expressed in some other doleful work. The third condition is less straightforward:

- (iii) the expressive object does not simply reveal what it expresses (as in mere physiognomic reading), but 'says' or 'communicates' it - where this is not just a metaphor for the agent's saying or communicating something with it.

Taylor gives two different kinds of example to illustrate this dichotomy. First, artistic; I may read a friend's tiredness or tension from an example of his sketches with which I am familiar, yet it is not this, but rather, say, a 'serene vision of things' which the sketch truly expresses.¹ Second, and more pertinently, the example of communicating friendliness, in two different ways: (a) by means of an open, welcoming smile, (b) in conditions of mutual marauding on the frontier, by my refraining from attacking your compound. In the first case, the smile expressing friendliness and openness is not just a matter of the agent's making it with the intention of communicating those things - the smile itself communicates friendliness, rather than just allowing the agent's friendliness to be seen, which is all that the behaviour in the second example does. Taylor writes:

... the smile's link to manifesting openness is anterior to the framing of all intentions to communicate anything. [It] plays a crucial role ontogenetically in our being able to enter into communication in the first place. Its expressing friendliness cannot thus just be a matter of its being deployed with that intent; although of course a smile may go dead, become a

¹ Why is it that tension can be 'read' from the sketch but only 'inferred' in the example of the friend's trembling hands earlier (p. 222)? We'll return to this matter after the exposition.

grimace, or look sinister, if the intent is absent (ibid. p. 78).

He concludes by noting that just as what it is for an expressive object Y to express X cannot be reduced to its offering a physiognomic reading of X, nor to its being created / uttered with the intent to communicate X, so it cannot be reduced to the sum of these things. I may work all night, so as to look tired, so as to get you to treat me better - but this doesn't make my tired expression into an expression of tiredness in the required, strong sense.

Well, this account of expression does seem to be on the right lines for present purposes. It is rough, and, as has on occasion been noted, needs clarifying - a task which will be imperfectly carried out here. Nevertheless, it is hoped that the advocacy of the treatment of avowals as expressive in Taylor's sense will be sufficient to persuade the reader that it is this fact which explains the Famous Five. Let us return to the extract on the friendly smile quoted above. We are finally in a position to tie together several ideas that have hitherto been too aimlessly floated.¹ The key seems to be that it is not that avowals are not uttered with the intention to communicate joy, pain, sadness, etc., but that characterising their role in such terms is insufficient. They may indeed be uttered with such an intention, as in the case where, say, I choose to let the medical practitioner know that I'm getting fed up with the discomfort occasioned by his ministrations by saying (improbably) 'I am in pain' (or, more likely, 'That hurts'). But the effectiveness of such utterances is parasitic on their more fundamental role as expressions of avowable states anterior to the framing of any intention to communicate.

This fact is connected with the positive presumptive aspect of avowals. Let us elaborate on this aspect. It may be suggested that it is not the mere utterance of 'I am in pain' that is authoritative, but the sincere utterance. It was in fact this suggestion that prompted the attempted resolution of p. 156, and we can now see why

¹ In the discussion of Wittgenstein's expressive account - pp. 189-94.

that was wrong. For what we have is not conditional authority (we make the proviso that the subject must be sincere) but prima facie authority defeasible in a certain way. That is, the provisos (sincerity, understanding, lack of self-deception etc.) on the avowal for it to guarantee the 3rd-person ascription are 'positive presumptive' - we assume they are met unless it is shown otherwise. And with regard to sincerity, what we are presuming is the absence of an intention (the intention to deceive). This is not to say that the intention to communicate, because it is not tested for, is therefore absent; simply that its presence has nothing essentially to do with why we regard an avowal as authoritative. The reason why the provisos of an avowal are positively presumptive is connected with the explanation of the authority of avowals itself; the common source for these features lies in the role of avowals as expressive antecedents of communication-intention.

What account of avowals is yielded by Taylor's three conditions on expression? First, we don't infer from the linguistic behaviour of avowal to the avowable state. This is clearly correct. (It is connected to the point about positive presumption; I do not check, before my 3rd-personal ascription to an avowing subject, that the provisos are satisfied). Second, that which an avowal expresses can be manifest only in an avowal or in some other piece of expressive, non-linguistic behaviour. Finally, the avowal itself 'says' or 'communicates' that which it expresses in a manner over and above that of the agent saying or communicating something with it.

(ii) Amending Taylor's account

This having been said, it may be questioned whether the avowal offers a 'physiognomic reading', even though condition (i) is satisfied in terms of non-inference. It may seem absurd to say that I 'read' off your pain from the 'physiognomy' of your utterance 'I am in pain'. But perhaps the term 'reading' misleads us here; all that is meant is that the expressive object presents us with a characteristic appearance which we see or recognize as an expression of pain, or whatever. And the characteristic tone of 'I am in pain' or 'I'm depressed' may well do just that, although this is combined

with the sense of the utterance in our recognition of the state. Avowals are peculiarly prone to have this aspect of tone which will be of significance too in explaining how they fulfil Taylor's condition (iii). The very words themselves must 'say' or 'communicate' something over and above the agent's saying or communicating something with them. It must be no accident that two senses of 'expression' come together here; the idea of an avowal as 'expressive' of a state and the idea of 'expression' as meaning 'with feeling'. The way we communicate in speech how we are feeling (in the widest sense of 'feeling') has something essentially to do with the feeling or tone that is apparent in our words. As was noted in the Wittgenstein passage already discussed, it is an inadequate model to think simply in terms of a subject intending to convey a thought by means of words, for the words themselves communicate his state by means of their tone.¹

It may be helpful to amend Taylor's account, and in fact simplify it, since his case for a distinction between objects which pass condition (i) but fail condition (ii) looks unconvincing. It is not clear that his example of the 'reading' of an impending fall in a building is non-inferential, since even where no inference is made, the grounds may yet be inferential. Problems of tense complicate this example, and we may take a simpler one. The molten steel's hotness is manifest in its seething, bluish appearance (I am viewing from a distance). I make no inference here from the appearance to the hotness; yet it seems that I could spell out a 'non-tautological' generalisation which would imply that we do have a 'physiognomic reading' here in Taylor's sense. ('Certain metals, when very hot, present a seething, bluish appearance, say.) There will thus be two senses of 'looks Ø':

(i) where looks Ø = presents an appearance reliably
inductively associated with (passing
independent tests for) being Ø [the
steel case]

(ii) where looks Ø ≠ " " [expression]

¹ See above pp. 196-97.

Rather than having three conditions which genuine cases of expression must satisfy, it seems better to have one condition for which Taylor's conditions (ii) and (iii) provide supplementary features. This condition is:

A manifestation counts as an expression of a state in a certain context just when the question 'Why does the state have that kind of manifestation?' is foolish or has no answer.

This formulation provides a test for the so-called 'logical connection' present in genuine cases of expression, and absent in those cases where there is an (inductive) generalisation to answer the question above. It also gets us round the problem of cases where no inference is actually made, yet where one could be furnished. The question 'Why does the state have that kind of manifestation?' is deliberately general. For it could presumably be asked, in the case of a subject in pain, 'Why does this state manifest itself in the odd bit of pain-behaviour (wincing, say) rather than in another way (by means of an avowal)?' - and an answer would be: 'stoicism'.

Let us say the subject's anxiety is manifested not in an anxious appearance, but in displacement activity. He picks a quarrel with a friend over something quite trivial, say. The question 'Why does the state of anxiety have that kind of manifestation?' is not foolish or unanswerable - since there is an inductive generalisation to answer it, of the form 'People, when anxious, often engage in apparently pointless or unmotivated behaviour as a means of masking the source of worry instead of effectively dealing with it'. One could make the discovery that anxiety can have such behaviour as a consequence; but one could not discover that pain could have as a consequence the subject writhing on the ground clutching his leg, say.

(iii) Defeasibility of mental-episode ascriptions

Before discussing the effect of using the new formulation of 'expression', it is worth saying something about a paradox adverted to when it was said above that our question is 'deliberately

general'. This point needs to be re-expressed - and the notion of a mental 'episode' clarified (involving some further correction of Rorty's ideas quoted on p. 163). The 'paradox' is the following:

(A) If a state is an episode, a subject can pass in and out of it without affecting the possibility of any longer-term, dispositional states obtaining.

(B) Ascriptions of mental episodes can be defeated in the light of subsequent behaviour.

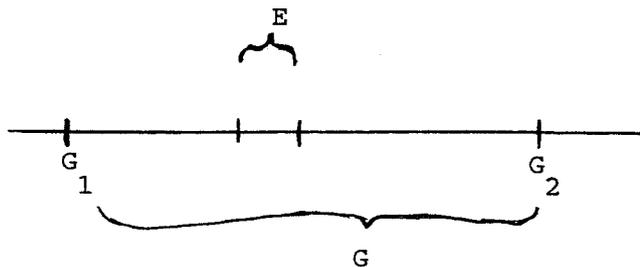
Rorty, in the passage from his [1970] quoted above (p. 163), has glided over the distinction here. He wrote (to re-formulate without altering his drift) that ascriptions of dispositions 'are implicit predictions of future behaviour, predictions which may be falsified'. In contrast, ascriptions of mental episodes may be falsified by later behaviour; someone displays pain-behaviour but later admits that he was shamming, tells us why, etc. How then are episodes compatible with any range of future behaviour? The question is evidently more subtle than was first supposed - and principle (A) is clearly in need of refinement. For present purposes however (and readers should know by now what this phrase means!), it is enough to indicate that episodic ascriptions are compatible with any range of future behaviour in that they are not predictions about it - though it is wrong to imply on the basis of this, as Rorty does, that they cannot be falsified by it. (If he had thought about it, he would have been more careful in what he said.) An expressive episode is presumed to be sincere - and so when I say 'He is in pain', I am not making a prediction that the subject will not subsequently disavow his apparent pain and admit he was shamming. This point is partly that already emphasized - the 'positive presumptive' aspect of avowals. It should be noted that any mental episode may be 'expressed' - but in some cases the expression is limited to linguistic behaviour (e.g. the cases of mental images, after-images, certain kinds of noises in the head, etc. etc.). And thus far, clearly, the linguistic expression is by means of 'paradigm' avowals - we will come on to 'qualified' avowals shortly.

The converse of the latter presumption is the fact that behaviour is expressive only relative to a context. If this were

not so, ascriptions of mental episodes could not be defeated.
Consider the two interpretations of our question above (p. 227):

- (1) Why does the subject express his state that way (on this occasion)?
- (2) Why is that kind of behaviour ever an expression of that kind of state?

The second question provides the formulation required for our test of 'expression'; its foolishness, in the case of expressive behaviour, is consistent with a particular piece of apparently expressive behaviour not being an expression of anything (but being stimulated, shammed, suggested under hypnosis). This is because behaviour (E) (below) expresses a state relative to a context or span of behaviour (G) - which may not, at the time of the behavioural episode, be fully disclosed:



Say the subject is in pain - then G expresses that state, and so do particular episodes such as E. But suppose the subject is not in pain - then neither G nor some apparently pain-expressing episode E in fact express pain. Perhaps the point is being laboured - but, to conclude, what is defeated is the claim that E is an expression of pain. (We would not say 'that's an expression of pain, but he's not in pain'.)

(iv) Consequences of the amended account of 'expression'

The upshot of the proposed amendment of Taylor's account is that it is simplified. There seems no reason to suppose that there are 'physiognomic readings' of non-expressive manifestations of states, and we are provided instead with a straightforward

distinction between expressive manifestations which are 'read' physiognomically and non-expressive manifestations which are 'read' (if you like) inferentially. The points about positive presumption and the role of avowals as expressive antecedents of communication-intention remain as stated. It is not possible in the space available at present to clarify much further the distinction between manifestation of a state by its mere causal consequences, and expressive manifestation. There is a grey area here which may cause concern. As has been noted, Taylor's comments seem inconsistent. He says that I may, on the basis of my experience of you, infer from the fact that your hand is trembling to the fact that you are agitated. But then he claims that I may gain a physiognomic reading of a friend's tiredness or tension from a sketch of his (without this being what the sketch genuinely expresses). Well, we've dropped the distinction between expression and mere physiognomic reading, and so it looks now as if the latter example belongs with the former on the inferential side of the dichotomy. But anxiety or tension is itself a tricky example for present purposes. Some people exhibit a flustered appearance which betokens no real inner anxiety (it is more like an act to help them get what they want, say), whilst the outwardly calm appearance of others belies the churning emotions inside. (The resemblance to the manifestations of pain in the whinger and the stoic is superficial here.) We thus find it hard to say what is 'read' physiognomically and what inferred when it comes to states of anxiety in others. But in proportion as this difficulty seems intractable, so we must regard anxiety as a dispositional, non-avowable state - like depression. All genuinely avowable states must express themselves in manifestations to which our question has a foolish answer or none at all - and also, in most cases, be inferrable from manifestations to which the question does have a sensible answer. It is in the nature of avowable states to be expressed - that is our central contention.

The expressive in fact seems to occupy a middle ground between the mere symptom and the fully intentional communication of a state. We are unhappy to regard a state as being 'expressed' by something that just happens to the agent as a result of his being in it, and instead see the state as 'revealed' or 'betrayed'. But equally (as the problem of avowals testifies) we do not like to view the deliberated verbal report as an expression either. The mediation of the

manifestation by communication-intention to the exclusion of the 'communication' the words themselves achieve by means of their tone seems to rule out some avowal-like statements as examples of genuine expression. (Hence the 'non-avowal' uses of certain avowals.)

(v) Dealing with 'qualified' avowals

It will not have escaped attention that following the highly selective accounts of different kinds of avowal on pp. 148-79, the problem of avowal has been discussed without any attempt to differentiate in the same way. In fact, such is the multifarious nature of the class of avowals that a proper attempt to relate the preceding resolution to its idiosyncrasies is quite beyond the scope of the present work. Instead, a gesture in this direction involving further generalisation will be attempted.

Distinction has already been made, in the class of avowable states, between those which are episodic and which issue in 'paradigm' avowals, and those which tend towards the dispositional, non-avowable psychological states and which issue in 'qualified' avowals.

Clearly, the preceding account of what it means to say that avowals are expressive, works best for 'paradigm' avowals. It is a difficult question how far it should be applied to the 'qualified' avowals at the other end of the continuum. For the expression of intentional states must be viewed as being constrained by the operation of some principle (or principles) of rationality - which generates the well-known sort of holism mentioned above. This fact must lead one to question the value of saying that 'qualified' avowals are expressive in anything like the same way as are 'paradigm' avowals.

It might be thought, then, that 'qualified' avowals are somewhat of an embarrassment to the present project. But in fact certain features limiting the way they exhibit the classic Famous Five have already been discussed - amounting in fact to a modicum of corrigibility (see pp. 152-56). So perhaps the fact that 'qualified' avowals are not straightforwardly expressive, should come as no

surprise. Let us examine this latter feature more closely, though. There seem to be three options here - none particularly attractive, it must be said:

- (i) There is both linguistic and non-linguistic expression of intentional states.
- (ii) There is linguistic expression of intentional states (viz. 'qualified' avowals) but the non-linguistic behaviour is not genuinely expressive.
- (iii) Neither linguistic or non-linguistic behaviour can genuinely express intentional states.

The first thesis would be endorsed by Charles Taylor, who in his [1979] is concerned, inter alia, with showing that 'the natural expression of wanting is trying to get'. The idea that there is natural expression of a state does in fact seem to go hand-in-hand with the idea that the relevant avowal is expressive too - which makes thesis (ii) such an uncomfortable one. If one believes, say, that the dog's digging for the bone is a natural expression of his wanting the bone, then one will be favourably disposed to thesis (i). However, as was indicated above, the existence of holistic constraints on one's interpretation of human behaviour leads one to say that wanting is only conditionally given expression by some appropriate action. Only if I believe that drinking that glass of water will satisfy my desire for liquid refreshment, will my desire 'find expression' in my drinking it. This fact opens the possibility of a sensible answer in this case to our question 'Why does the state have that kind of manifestation?' (Viz. 'Because subjects reason that drinking water will quench their thirst'.) This in turn seems to create a divergence between the expressive avowal ('I am thirsty' - no sensible answer to our question) and the non-expressive behaviour. (It is because the avowal seems to pass the test for expression that thesis (iii) is the least plausible.)

The problem of 'qualified' avowals is not going to affect the plausibility of the expressive resolution to any marked degree, and so the question will not be treated in the depth it requires. However, it is worth saying that there is room for manoeuvre over

the expressive / non-expressive status of '"qualified" avowal behaviour'. For whether the behaviour passes the test for expression depends on how it is described. Is there not some formulation that will ensure the answer to our question is foolish? E.g. 'The state of being thirsty manifests itself in the subject trying to get drinkable, non-toxic liquid'.

The expressive resolution of the problem of avowals constitutes a conceptual explanation (as opposed to an institutional one - see p. 207 above). However, it is not the only possible such explanation for the features of 'qualified' avowals. If the latter are ultimately judged to fail the test for expression, they need not thereby be seen as reports. Rather, one may say the following. The reason why, in deciding whether p, I am also deciding whether I believe that p, is that, as suggested, my beliefs are not simply 'given' as are sensations and (in a certain range of cases) moods; it may not be a determinate matter whether, at some time t, I believe that p (and similarly for other propositional attitudes). The avowal of belief may best be seen as a decision rather than as a report - or as the reiteration of a decision in some cases. (The metaphor 'making up one's mind' is one to be taken seriously.) 3rd-personal belief-ascription depends on the disposition to verbal expression by the subject in a way that pain-ascription does not; the avowal of belief is an integral part of the process of belief-formation in a way that the avowal of pain is not in the 'process' of 'pain-formation'.

(vi) Avowals and 'pseudo'-avowals

Could there be a subject of experience that did not make avowals (that is, authoritative utterances expressing states of itself)? Only if solipsism were correct. For if avowals were authoritative because of some 'look within', there would have to be the coherent possibility that such introspection never yielded anything in the way of publicly-observable behaviour. They would then be 'pseudo'-avowals - mere reports on mental states, their attachment to which would be contingent (i.e. mental states would not need to be avowable or 'pseudo'-avowable). But the idea of a subject of experience which made only 'pseudo'-avowals in this

caricatural way is a nonsense. Its detachment from its own experience would be more profound even than its detachment (via the inferential model of 3rd-personal ascription) from the experience of others. If a subject is prone to states of consciousness he must be regarded as having the capacity to express those states; and their linguistic expression will be by means of avowals.

APPENDIX

ELIZABETH ANSCOMBE UND DER BANGOR KREIS

An account of a conversation between the author and Professor Anscombe.

10 July 1983 The foyer, Neuadd Rathbone, Coleg Prifysgol Gogledd Cymru, Bangor, Gwynedd.

Lunch-time: those present - G.E.M. Anscombe and A.J. Hamilton (Miss Anscombe is no longer wearing a name-tag pinned on to the back of her shirt).

GEMA I want a pint of beer

* * *

The bar, as above. It is rather noisy.

AJH I think that a lot of people don't seem to have fully understood your paper, 'The First Person', or at least the positive account you give of self-consciousness.

GEMA Oh, only Jenny Teichman has told me she didn't understand it ... maybe that's because she knows me as a friend ...
... Jimmy Altham wrote a paper on it, you know, it's in the collection, Intention and Intentionality.

AJH Yes, I've looked at that.

GEMA What will you have?

AJH Yes, I have read it.

GEMA No, what will you have to drink?

AJH Oh, a grapefruit juice please ... that's very kind of you, Professor Anscombe ...

* * *

In the foyer, as above.

AJH Do you still believe that 'I' is not a referring expression?

GEMA Oh yes.

AJH Perhaps I could ask you first about the consequences of your view. I have heard it said that it becomes a problem what happens to the truth-value links between 1st-person utterances in the present tense, and other utterances. For instance, if someone says, 'I am in pain', someone else can say of him 'He is in pain' ... Some people say that reference just consists in there being these truth-value links.

GEMA Well if that's so, 'I' does refer! ... I think it would be better to ask, how can one infer from a sentence containing a non-referring expression to one containing, in the relevant place, a referring one ... I'm thinking here of the stuff by Davidson ... I can infer from 'I stabbed him with a knife' to 'I stabbed him', but how is this done? ... Tony Kenny started this off in his book Action, Emotion and Will ...

[The content of Miss Anscombe's remarks tails off into indeterminacy as her interlocutor loses the drift.]

... perhaps we ought to have lunch now ...

* * *

The refectory, as above: those present - G.E.M.
Anscombe, P.T. Geach and A.J. Hamilton.

AJH In your article, you give explanations of the meaning of 'I' which you find unsatisfactory, because of their circularity. I'm thinking of "'I" is the word each one

uses knowingly and intentionally to refer to himself'. I realise this is circular, but could it not serve as an introductory explanation? ... I mean, it's not totally uninformative.

GEMA Well ... imagine this story. A character is looking for the heir to a fortune. He's given some information about him, he then finds out when he was born (this could be a laborious process), then where he lives, and so on. He goes to the house, knocks on the door, but the person he's looking for, John Smith let's say, isn't in. Then ... it dawns on him, 'this is my house! It is myself who is John Smith, the heir!' ... One could write a novel about this, a Borges sort of novel ... What would it be like to arrive at the realisation that this person, so fully described, but whom I haven't actually met, is myself! Extraordinary! ... I remember, when we were visiting --- they had houses with numbers in two streets ... you know, there was a garden on one side, and that was the usual entrance, and a garden on the other ... mail might be addressed to the number and street on the other side, and returned marked 'Undelivered' ... You know the 'Thou Art the Man' stories ...

AJH No ...

GEMA Oh, like Naboth's vineyard. The king takes it away from him ... The king consults the prophet, who tells him a story about the man who stole the sheep ... The king says, 'This man must be brought to justice! It is terrible! How might we find him?' And the prophet says, 'Thou are the man' ...

AJH But you still feel that the circularity of the account of 'I' you gave is a decisive objection?

GEMA Well, what does John Smith now know when he realises 'It's me who's the John Smith who ... '?

AJH Maybe the conditions on the use of 'I' can't be spelled

out. Maybe there's nothing there that can be spelled out.

GEMA Maybe.

AJH Do you think that 'I' is dispensable from the language, that we could say just the same things we presently say, without it?

GEMA Oh, of course! Latin has no 1st-person pronoun in common use, just a 1st-person inflection of the verb ... It's always useful to look at other languages you know - people writing in this area don't do this ... because usually they only know one language ... Descartes says 'Cogito ergo sum', not 'Ego cogito ergo ego sum' ... Then he says 'Quid est iste ego cogitans?'¹ ... is that it? Peter dear, does Descartes say 'Quid est iste ego cogitans?'?

PTG Is it [long quotation]?

GEMA Oh no, it's not that at all.

PTG Well, I don't know dear ...

AJH I only know one language ...
... So you hold that 'amo' doesn't serve to identify anyone?

GEMA No.

AJH Doesn't ... um ... 'amat' identify someone?

GEMA No.

AJH Hmm ... Strawson says in The Bounds of Sense that when I use 'I', I identify myself for others.²

¹'What is this thinking I?' - Ed.

²The relevant remark is in fact in Individuals p. 100 - Ed.

GEMA It's not by means of 'I' that I identify myself to others, but merely by speaking ... Imagine a lot of deep holes, and I fall down one of them ... I cry 'Help!' Someone asks 'Who's there?' "It's I", I say. Well, it's not 'I' that's doing the identifying ... rather, the tone of voice, accent, etc.

AJH I asked you earlier about the dispensability of 'I'. In the Philosophical Remarks, Wittgenstein talks of 'alternative techniques of representation', which do not employ the 1st-person pronoun ... 'There is pain here', etc.

GEMA Yes.

AJH Others have talked of dispensing with 'I' and preserving the expressive power of the language, of saying just what we can say now, using 'I'. Wittgenstein goes on to say of the alternative techniques of representation, that they preserve what is 'essential to the facts' ... what do you make of this notion - it seems very Tractarian ...

GEMA Yes, it would be ... You'd need to look at something later, the Philosophical Grammar ...

AJH Oh ...
... David Bell claims that in passages like the one from the Remarks, Wittgenstein is espousing a kind of solipsism ... rather like Strawson's 'No-ownership view' in Individuals.

GEMA I don't think the no-ownership view is solipsism ... There is that passage about the 'visual room' in the Investigations¹. The 'visual room' is one that has no owner ... the idea of the 'geometric eye' [Anscombe points to between her eyes] is implicit here ... I never understood this passage until I came to write 'The Intentionality of Sensation' ... about non-existent objects of

¹ Sections 398-400 - Ed.

perception, like in that experiment when they flash up FIN and RED alternately very rapidly and you seem to see FRIEND, but it isn't really there ... what kinds of object are these? ...

PTG Would you like coffee, dear?

GEMA Oh, yes please ...

AJH There is a continuing strain of solipsism in Wittgenstein's work, though, isn't there? He does say in a passage in the Lectures edited by Alice Ambrose, that what the solipsist wants is not a language where the ego has a monopoly, but where it drops out of the picture.

GEMA Well, of course, in the Tractatus, he does say that solipsism ultimately coincides with realism, you can develop it that way [TLP 5.64 - Ed.].

AJH I find it strange that while Wittgenstein devotes much discussion in the Investigations to the alleged 'mental objects', which are idle, which must drop out of the picture, he doesn't seem to discuss the subject any more in the same way, in the way he does in the middle period writings. Why is it that there is so little discussion of the self in the Investigations?

GEMA [with exasperation] How could a man whose native language has no expression for the self possibly discuss it! You really must stop using this expression!

AJH [obsequiously] Er, yes, ... um ... I should really have said 'the 1st-person'.

GEMA The 'I'.

PTG I'd like the keys dear. I must go and do some work this afternoon.

[GEMA gives the keys to PTG and he departs]

GEMA Yes, so must I ... Anyway, I did object to the way McGuinness, in the Tractatus, translated 'das Ich' as 'the self'.

* * *

In the foyer: those present - G.E.M. Anscombe and A.J. Hamilton. Immediately after the above.

AJH I wasn't very clear about your discussion of 'A'- and 'I'-users in 'The First Person'.

GEMA No, I'm not happy with it. A lot of what I wrote on it didn't appear in the published article.

AJH Is the idea that with 'A', it's just like 'I' except that identification makes sense?

GEMA Well, my aim here was this. People say 'I' is a name. Then let's treat it like one and call it 'A'. 'A' is the name everyone uses [intentionally? - Ed.] to refer to himself. Everyone has 'A' stamped on his wrist. Then, imagine someone is wounded in the leg. He has to ...

AJH Trace a path ...

GEMA Yes, trace a path from the wound to the mark on his wrist. Then he can say 'A' is wounded. The trouble is, how could someone, for instance, trace a path from his wound to the 'A' stamped on his wrist without knowing 'this is my body'?

AJH They would have to be self-conscious, anyway, then?

GEMA Conscious, at least. Remember that example of the bird 'dropping a wing' ... Pretending, but not knowing its pretending. Anyway, I do recommend you follow up this line of thought. Maybe it leads somewhere, maybe not.

AJH A kind of language for Baldy ...

GEMA Oh yes, William James is so good, isn't he!

AJH I particularly liked the bit where he writes 'X, alias
"Baldy"' ...

GEMA Yes! ... Well, thank you. Goodbye.

AJH Yes, thank you very much, Professor Anscombe.

REFERENCES

- Altham, J. [1979] 'Indirect Reflexives and Indirect Speech', in Intention and Intentionality: Essays in Honour of G.E.M. Anscombe, ed. C. Diamond and J. Teichman.
- Ambrose, A. [1979] Wittgenstein's Lectures, Cambridge 1932-1935, Oxford: University Press.
- Anscombe, G. [1975] 'The First Person', in Mind and Language ed. S. Guttenplan, also in G. Anscombe [1981], Collected Philosophical Papers Vol. II: Metaphysics and the Philosophy of Mind, Oxford: Blackwell.
- [unpublished] 'Propositions'.
- Anscombe, G. and Geach, P., eds. [1970] Descartes: Philosophical Writings, London: Nelson.
- Armstrong, D. [1963] 'Is Introspective Knowledge Incorrigible?', Philosophical Review, Vol. 72.
- Austen, J. [various] Emma, various editions.
- Austin, J. [1962] Sense and Sensibilia, Oxford: University Press.
- [1962] How To Do Things With Words, Oxford: University Press.
- Ayer, A., ed. [1959] Logical Positivism, Glencoe: The Free Press.
- Bell, D. [unpublished] 'Solipsism'
- Blackburn, S. [1984] 'The Individual Strikes Back', Synthese Vol. 58.
- [1984] Spreading the Word, Oxford: University Press.
- [1986] 'Thoughts Without Representation', Proceedings of the Aristotelian Society, Supplementary Volume.
- Buck, R. [1962] 'Non-Other Minds', in Analytical Philosophy, 1st series, ed. R. Butler.
- Butler, R. [1962] Analytical Philosophy, 1st series, Oxford: Blackwell.
- Carnap, R. [1959] 'Psychology in Physical Language', in Logical Positivism ed. A. Ayer.
- Diamond, C. and Teichman, J., eds [1979] Intention and Intentionality: Essays in Honour of G.E.M. Anscombe, Brighton: Harvester.
- Dickens, C. [various] Hard Times, various editions.

- Dummett, M. [1976] 'What is a Theory of Meaning? (II)' in Truth and Meaning ed. G. Evans and J. McDowell.
- [1978] Truth and Other Enigmas, London: Duckworth.
- [1981] Frege: Philosophy of Language (2nd edition), London: Duckworth.
- Evans, G. [1973] 'The Causal Theory of Names', Proceedings of the Aristotelian Society, Supplementary Volume.
- [1982] The Varieties of Reference, Oxford: Clarendon.
- Evans, G. and [1976] Truth and Meaning, Oxford: Clarendon.
McDowell, J., eds.
- Flanagan, O. [1984] The Science of the Mind, Cambridge, Massachusetts: MIT.
- Geach, P. [1965] 'Assertion', Philosophical Review, Vol. 74.
- Ginet, C. [1975] Knowledge, Perception, and Memory, Dordrecht: Reidel.
- Guttenplan, S., [1975] Mind and Language, Oxford: Clarendon.
ed.
- Hacker, P. [1972] Insight and Illusion, Oxford: Clarendon.
- [1986] Insight and Illusion (revised 2nd edition), Oxford: Clarendon.
- Hamilton, A. [1986] 'Bill Evans: The Village Vanguard Sessions', Wire (January).
- [forthcoming] review of P. Hacker: Insight and Illusion (2nd edition), Philosophical Quarterly.
- Hare, R. [1952] The Language of Morals, Oxford: University Press.
- Harrison, R., [1979] Rational Action, Cambridge: University
ed. Press.
- Holtzmann, S. and [1981] Wittgenstein: To Follow a Rule, London:
Leich, C., eds. Routledge.
- Hoxha, E. [1986] The Artful Albanian: Memoirs of Enver Hoxha, London: Chatto and Windus.
- James, W. [1890] Principles of Psychology, London: Macmillan.
- Janaway, C. [1984] 'The Subject and the Objective Order', Proceedings of the Aristotelian Society, Supplementary Volume.

- Kant, I. [various] Critique of Pure Reason, various editions.
- Kenny, A. [1968] Descartes: A Study of His Philosophy, New York: Random House.
- [1973] The Anatomy of the Soul, Oxford: Clarendon.
- [1979] 'The First Person', in Intention and Intentionality: Essays in Honour of G.E.M. Anscombe, ed. C. Diamond and J. Teichman.
- Kripke, S. [1973] 'The Causal Theory of Names', Proceedings of the Aristotelian Society, Supplementary Volume.
- [1982] Wittgenstein on Rules and Private Language, Oxford: Blackwell.
- Luckhardt, C. [1983] 'Wittgenstein and Behaviourism', Synthese Vol. 56.
- McGinn, C. [1983] The Subjective View, Oxford: Clarendon.
- MacIntyre, A. [1956] The Unconscious, London: Routledge.
- Malcolm, N. [1968] 'Wittgenstein's Philosophical Investigations' in Wittgenstein: The Philosophical Investigations ed. G. Pitcher.
- Moore, G. [1955] 'Wittgenstein's Lectures in 1930-33' Mind, Vol. 64.
- Nagel, T. [1970] The Possibility of Altruism, Oxford: Clarendon.
- Nerlich, G. [1967] '"If You Can't Be Wrong, Then You Can't Be Right"', Philosophical Quarterly, Vol. 72.
- Noonan, H. [1979] 'Identity and the First Person', in Intention and Intentionality: Essays in Honour of G.E.M. Anscombe, ed. C. Diamond and J. Teichman.
- Nozick, R. [1981] Philosophical Explanations, Oxford: Clarendon.
- Parfit, D. [1971] 'Personal Identity', in Philosophical Review, Vol. 80.
- [1984] Reasons and Persons, Oxford: University Press.
- Peacocke, C. [1983] Sense and Content, Oxford: Clarendon.
- Pears, D. [1975] 'Wittgenstein's Treatment of Solipsism in the Tractatus', in his Questions in the Philosophy of Mind.

- [1975] Questions in the Philosophy of Mind, London: Duckworth.
- Perry, J. [1979] 'The Problem of the Essential Indexical', Nous Vol. 13.
- [1986] 'Thoughts Without Representation', Proceedings of the Aristotelian Society, Supplementary Volume.
- Pettit, P. and McDowell, J., eds [1986] Subject, Thought and Context, Oxford: University Press.
- Pitcher, G. [1971] A Theory of Perception, Princeton: University Press.
- Pitcher, G., ed. [1968] Wittgenstein: The Philosophical Investigations, London: Macmillan.
- Platts, M. [1979] Ways of Meaning: An Introduction to a Philosophy of Language, London: Routledge.
- Price, H. [1983] '"Could A Question Be True?": Assent and the Basis of Meaning', Philosophical Quarterly, Vol. 33.
- Richards, W. [1984] 'Self-Consciousness and Agency', Synthese, Vol. 61.
- Rorty, R. [1970] 'Incorrigibility as the Mark of the Mental', Journal of Philosophy, Vol. 67.
- Ryle, G. [1949] The Concept of Mind, London: Hutchinson.
- Sacks, O. [1985] The Man Who Mistook His Wife For A Hat, London: Duckworth.
- Schilpp, P., ed. [1952] The Philosophy of G.E. Moore, New York: Tudor.
- Shoemaker, S. [1968] 'Self-Reference and Self-Awareness', in Journal of Philosophy, Vol. 65.
- van Straaten, Z. [1980] Philosophical Subjects: Essays Presented to P.F. Strawson, Oxford: Clarendon.
- Strawson, P. [1959] Individuals, London: Methuen.
- [1966] The Bounds of Sense, London: Methuen.
- [1971] Logico-Linguistic Papers, London: Methuen.
- [1980] 'P.F. Strawson Replies', in Philosophical Subjects: Essays Presented to P.F. Strawson, ed. van Straaten.

- Taylor, C. [1979] 'Action As Expression' in Intention and Intentionality: Essays in Honour of G.E.M. Anscombe, ed. C. Diamond and J. Teichman.
- Waismann, F. [1979] Wittgenstein and the Vienna Circle: conversations recorded by Friedrich Waismann, Oxford: Blackwell.
- Wittgenstein, L. see below
- Wollheim, R. [1979] 'Wish-fulfilment' in Rational Action, ed. R. Harrison.
- Woodfield, A., [1982] Thought and Object, Oxford: University ed. Press.
- Wright, C. [1983] Frege's Conception of Numbers as Objects, Aberdeen: University Press.
- von Wright, G. [1971] Explanation and Understanding, London: Routledge.

Works by Wittgenstein are referred to by the following abbreviations:

BB	<u>The Blue and Brown Books</u>	[1960]
NB	<u>Notebooks 1914-1916</u>	[1961]
NFL	'Notes for Lectures on "Private Experience" and "Sense Data"'	[1968]
OC	<u>On Certainty</u>	[1969]
PG	<u>Philosophical Grammar</u>	[1969]
PI	<u>Philosophical Investigations</u>	[1958]
PR	<u>Philosophical Remarks</u>	[1964]
RPP	<u>Remarks on the Philosophy of Psychology</u>	[1980]
TLP	<u>Tractatus Logico-Philosophicus</u>	[1961]
Z	<u>Zettel</u>	[1981]

All are published by Oxford: Blackwell, except NFL (The Philosophical Review, Vol. 77) and TLP (London: Routledge).

References to fictitious works

- Ballard, Mr. "A Queer Memory Phenomenon" and Other Recollections
- X, alias 'Baldy' Adventures, serialised in Baltimore Oriole, 1890.

Regrettably neither of these works exists.