P H Y S I C A L I S M

A N D

P R I V A C Y

(A thesis submitted for
the D. Phil. degree.)

By

Daniel Altmann



Michaelmas Term, 1977 [i.e 1978] St. Edmund Hall

"If a body with clothes an' a head an' a face an' a hat doesn't make a person, I don't know what does."

Richmal Crompton (through William Brown)

ABSTRACT

The main objective of this thesis is to indicate an approach through which ontological dualism of mind and body may be collapsed - to show that (and how) the intuitive resistance to ontological monism is confused.

A minor objective, much less extensively argued for, is to indicate that once we can accept that there is no logical obstacle to the view that we are purely physical, then human mentality poses no logical obstacle to the scientific accommodation of human beings - to physicalism.

It is argued that we already have, in Wittgenstein's Private Language Argument, a very powerful argument suggesting that ontological dualism is logically unsound. The assessment of Wittgenstein's impact on ontological dualism occupies the first major section of the thesis (Part 2).

But it is found that the logical force of Wittgenstein's argument, though successfully applicable against ontological dualism, does not prescribe the sort of monistic account we should adopt. For there are two alternatives which satisfy Wittgenstein's argument, which is essentially an argument banishing necessarily "private" mental events. One alternative involves abandoning the view that we can be introspectively aware of mental events (essentially a behaviourist approach). The other alternative is to retain the possibility of introspective awareness of mental events, while maintaining that the latter are only contingently "private" (an approach with which the mind-brain identity theory may be aligned).

The first alternative is summarily rejected as being too counter-intuitive; and the remainder of the thesis explores the viability of the second alternative. But the latter alternative is also counter-intuitive: the dualist and the unconvinced materialist resist the suggestion that the mental events of which we may be introspectively aware could be neural events occurring in the brain or central nervous system.

In Part 3 the mind-brain identity theory is discussed. The strategy underlying this approach, as generally conceived, is found not only to be

unstable and ambivalent - straddling two rather different views - but also to be marked by a reluctance to engage sympathetically with the dualist's resistance.

In Part 4 an attempt is made to examine and undermine this resistance. It is found that this resistance is set in a context involving a confused form of realism; and the confusion is traced to a familiarly mistaken notion of perception, in which mental perceptual events are taken to mediate between a "mental subject" and the "external world".

The attempt to expose the confusions involved here, and to present a more satisfactory realism in a monistic setting, is supported by a formal-linguistic treatment of the relevant aspects of perception. In this formal account, which requires some elementary set-theoretical notions (in particular the notion of isomorphism), a designatory role is defined for perceptual events taken as syntactic entities in a certain sort of formal language. Through this formal treatment it is shown that for a rich enough (purely) physical structure there would be a "subjective dualism": essentially a symptom of the fact that for a physical structure to "perceive" a physical event, there would have to occur in it an unperceived (physical) event. And it is suggested that the dualist's resistance is based on a confusion in which what he takes to be a justification for ontological dualism can only be taken as a justification for "subjective dualism".

As a result of these considerations a modified form of the mind-brain identity theory is advocated, in which mental events for whose occurrence we can have introspective evidence are construed as unperceived (but not imperceptible) physical events which (it is hypothesised in the case of human beings) have neurophysiological descriptions.

In Part 5 this view is considered in a more general context. Also in this section there is an argument for a view hinted at earlier in the thesis, maintaining that the peculiarities of "mental discourse" pose no serious problem for physicalism.

Finally, two problems connected with the notions of the "unity" and

"simplicity" of mind are briefly mentioned, and an indication is given as to how they may be handled by the present account.

Length of thesis: approximately 75,000 words (including citations).

ACKNOWLEDGEMENTS

I am deeply indebted to my supervisor Mr. John McDowell for his invaluable and patiently given advice both on matters of form and of detail. He has saved me from making many embarrassing errors. Naturally I am entirely to blame for the faults which remain.

I wish to thank Mr. Paul Snowdon for his very helpful comments on some drafts of parts of this thesis, and for his encouragement generally. I have also been assisted through discussions of the mind-brain identity theory with Mr. J.M.Hinton; and through discussions of Wittgenstein with Mr. D.F.Pears, without whose advice I would not have attempted this project.

I am very grateful to my father for his interest and encouragement, and to Ms. Charlotte Humphrey for her untiring help during the final stages in the preparation of this thesis.

C O N T E N T S

PART 1	INTRODUCTION	1
1.1	Ontological Dualism	3
1.2	Approaches to Physicalist Monism	9
Notes to Part 1		13
PART 2	PRIVACY LANGUAGE AND PHYSICALISM	14
2.1	Two Stages	16
2.2	Reference to Private Entities: Stage I	18
2.3	Residual Privacy and the Ozma Problem	38
2.4	Stage I, Stage II, and Physicalism	43
2.5	The Possibility of P-Primacy	68
2.6	The Banishment of Private Mental Events and the Problem of Introspection	85
Notes to Part 2		93
PART 3	THE MIND-BRAIN IDENTITY THEORY	104
3.1	Problems of the General Strategy	105
3.2	The Identity Hypothesis and Leibniz's Principle	113
3.3	Smart's Analysis of Sensation Reports	116
3.4	Two Identity Theories	122
3.5	Incorrigibility and Imperceptibility	124
3.6	The Location Objection	130
3.7	Explanation and the Identity Theory	138
3.8	Kripke's Argument Against the Identity Theory	143
3.9	Davidson's Argument For the Identity Theory	149
3.10	The Identity Theory - So Far	156
Notes to Part 3		159
PART 4	SUBJECTIVE DUALISM AND ONTOLOGICAL MONISM	172
4.1	Monism Resisted	172
4.2	The Resistance Undermined	175
4.3	A Schema for the Metatheory of J-Languages	184

4.4 The J-Linguistic Treatment for Mental Perceptual Events 204

4.5 Subjective Dualism, Ontological Monism and the Augmented Identity Theory 227

Notes to Part 4 235

PART 5 THE AUGMENTED IDENTITY THEORY 238

5.1 The Right Distinction Wrongly Drawn 238

5.2 The Augmented Identity Theory and Physicalism 246

5.3 Problems of Unity and Divisibility 261

5.4 Conclusion 267

Notes to Part 5 270

REFERENCES 275

P A R T 1

INTRODUCTION

My purpose is to try to render the "mental life" of human beings compatible with physicalism. To do this I will isolate what appears to be a crucial obstacle which the mental seems to place in the way of physicalism, a difficulty which I will argue is inadequately dealt with by standard approaches to physicalism; and I will indicate an approach through which this difficulty may hopefully be dissolved. I will not try to prove that physicalism is a true doctrine, nor will I be particularly concerned to stress the advantages of a physicalist position. I will merely suggest how the mind may possibly be accommodated within a physicalist view.

I will take physicalism and materialism to be more or less the same doctrine, roughly the view that the world is composed only of the sort of entities postulated by and in principle susceptible to investigation and explanation by the physical sciences; these entities are of course physical entities. However, I will generally use 'physicalism' as the name of this doctrine only in so far as it is relevant to an account of human beings. For this context I will distinguish between a weak and a strong form of physicalism.

I will take physicalism in its weaker form as maintaining that there is in principle available a physicalist description and explanation (essentially, a description and explanation couched in scientific-theoretical terms) for every event occurring in a human being and in which a human being partakes. And I will take physicalism in its stronger form as claiming that the human mind presents no special logical obstacle to the scientific accommodation of human beings and their behaviour. (By a "special logical obstacle" I mean a logical obstacle which is

presented by human beings but not by physical structures in general.)

Clearly an ontological monism admitting just physical events is necessary for physicalism in its weaker form. But ontological monism may not be sufficient: for there may be grounds for supposing that even if we are purely physical structures, some physical events in us would in principle defy scientific accommodation. However, I will assume that there are no such grounds, and that ontological monism is necessary and sufficient for physicalism in its weaker form. (My justification for this assumption is partly that I don't know of any such grounds; and mainly that if we were purely physical, such grounds could not readily be interpreted as a problem which the human mind poses for science - and it is just that sort of problem with which I am concerned.)

It is by no means obvious that an ontological monism admitting just physical events is sufficient for the stronger form of physicalism (though again it is clearly necessary). It has been supposed, for instance, that certain features of the mental which cannot readily be taken to require an ontological dualism of mind and body, are nevertheless troublesome to physicalism and science. But I will argue that the only features of the mental which could obstruct physicalism in the stronger form would be those which prevented the collapse of an ontological dualism of mind and body; and that therefore the weaker form of physicalism is sufficient for the stronger form. So those features of the mental which seem to prevent ontological monism present the crucial obstacle to physicalism (in either form). And I will endorse the view that a meaning-irreducible dualism concerning the way our behaviour is described, which does not entail ontological dualism, in itself presents no logical difficulty for physicalism in the stronger form.

We may now briefly remind ourselves of the sort of intuitive pressure which seems to force us to an ontological dualism, and of some of the problems facing this position.[1]

1.1 Ontological Dualism

Descartes' distinction between mind and body was put in terms of two sorts of substances. We may find it more immediately intelligible to consider the supposed ontological dualism as involving two radically different sorts of event.

Many mental attributions cannot be taken as asserting the occurrence of mental events:- for example, 'A wants to be a composer', 'A believes in God'. On the other hand 'A had a yellow after-imaging', 'A had a stabbing pain', 'In a flash A conceived the solution to the problem' seem to assert the occurrence of events. A dualist might make the following remarks to justify his view that these events differ radically from physical events:

'We may distinguish a class of events which are in principle imperceptible - mental events - as opposed to physical events which are in principle perceptible.'

'Thus, when I have pains, kinaesthetic sensations, when trains of thought occur in me, and so on, I have direct access to or direct evidence for the occurrence of these events, whereas other people can only have indirect evidence for their occurrence. No one could perceive the mental event which occurs in me when I have a pain, for instance, no matter how closely he examines my body. Yet these events definitely occur in me.'

'And the same goes for mental perceptual events: for when I see the water boiling, or hear the water boiling, not only does there occur the event we describe as the water boiling, but also there occurs the perceptual event, my visual perceiving, (or my auditory perceiving). And such events are in principle imperceptible to others, who can only have indirect evidence for their occurrence, whereas I have direct access to them.'

'And in as much as mental events are conceived as being in principle imperceptible, they cannot be conceived to be spatially related to physical events - they cannot be conceived as having spatial positions - only as having temporal positions.'

The emphasis here on imperceptibility does not occur in Descartes' account, though it is implicit there. For some such distinction in terms of imperceptibility is central to ontological dualism, in that it is required by the claim that mental events are additional to the physical events occurring in the body. The notion of imperceptibility in this context is problematic, as I will mention in Part 3, where I discuss the mind-brain identity theory.[2] But we can easily see that what the dualist is getting at is that even though we may contingently not be able to perceive directly certain physical events, we can nevertheless conceive how such physical events engage causally with contingently perceptible events which may thereby provide indirect evidence for the occurrence of the former. But mental events to the dualist seem to have to be different even from contingently imperceptible (physical) events - in particular, different from and additional to the physical events occurring in the body - because he cannot conceive how mental events could engage causally with physical events. And in this way we can give preliminary significance to the dualist's claim that mental events are in principle imperceptible.[3]

So we may note that the supposed distinction between mental and physical events seems to rely on the view that a person somehow has direct access to his mental events - he is in principle in a privileged position with regard to his mental events - and other people in principle can only have indirect access to them. It is in this sense that we may regard the dualist as claiming that mental events are in principle private to one person.[4]

Exactly what is the nature of privileged access in the dualist's account - exactly what is contained in the claim that a person's mental events are in principle private to him - is not altogether clear. Various symptoms of privileged access have been suggested, some of which I will consider when I discuss the mind-brain identity theory. But one of the most frustrating aspects for the dualist involved in criticisms of his position is that the notion of privileged access is most readily accessible through various linguistic symptoms; but the critical treatment of these symptoms tends to lack both sympathy for the underlying conviction of the dualist, and the power to undermine his motivations. In Part 4 I will try to undermine the motivations which lead the dualist to suppose that a notion of logically privileged access is needed in connection with mental events. But in Part 2 I will take the notion of logically privileged access as primitive, and as far as possible discuss its consequences without trying to analyse the notion itself. We will find that Wittgenstein's arguments render such a notion incoherent.

For the moment I will not produce any further clarification of the contention that mental events are in principle private to one person, nor will I criticise the dualist on the grounds that he could not produce any clarification. But in order to assess the problematic consequences of privacy I will find it helpful to distinguish between the imperceptibility and incommunicability components of privacy. We have already seen how there seems required by the dualist a sense in which a person's mental events are in principle imperceptible to others. It seems likely also that the notion of logically privileged access as used by the dualist requires that there is something about a person's mental events which is in principle <u>incommunicable</u> by him to others. I will discuss this possibility in Part 2.

Now, one of the most important features which a dualist view

containing this sort of distinction between the mental and the physical shares with the Cartesian account is the type of realism which it implicitly maintains. This realist view, accompanied by a representative theory of perception, supposes that the mental perceptual events in a person convey information to him about the physical world external to him; and what allows this realist view to encourage ontological dualism is that it is implicitly supposed that each individual has direct access to a "private bubble" comprising his sensations, and so forth, and indirect access to the external world through his private sensations. Without this type of realism an ontological distinction between the mental and the physical cannot readily get off the ground. But a realist view of this sort has some rather serious problems:

(i) There are two difficulties which may be regarded as two sides of one coin, concerning how we communicate given the privacy of the mental, and concerning our relation with the external world. These difficulties also present themselves in two intimately related ways: as problems of meaning, and as epistemological problems. We may exemplify the sort of problems of meaning involved by the problem concerning whether we are to interpret expressions like 'the table before me' as referring to "groups of sensations" or to "objects in the external world" - neither interpretation is on the realist view under consideration without problems. And the epistemological problems centre around the question of how much, if anything, a person can know or infer about the external world on the basis of his sensations, and how such knowledge would compare with his knowledge of his private sensations.

Thus we have various views (such as Locke's) where a person's immediate evidence concerning his sensations is contrasted with his mediate evidence concerning the external world. Such views, as Wittgenstein suggests, tend to suppress serious difficulties concerning

communication and public meaning. (Of course, if the realist distinction between perceptions and the external world is abandoned, as in Berkeley's view, then whatever distinctions remain available in such a view, ontological dualism is not one of them.)

(ii) There are serious difficulties for this form of realism concerning the coherence of the notion of the "person", the subject who has direct access to "his" private bubble, "in whom" private mental events occur.

These two sorts of problem can be taken as applying to the internal coherence of ontological dualism and the realist view in which it is embedded. And the following difficulties involve problematic consequences of ontological dualism as soon as we allow its coherence:

(iii) (a) What is the relation between a person's mental events and his voluntary bodily movements? Descartes wanted a means by which certain mental events could influence or direct voluntary bodily movements, in order to allow room for the notion of freedom of the will.

(b) What is the relation between the physical events perceived by a person and the mental perceptual events in that person? Descartes wanted to be able to say that there was some relation:

> ... from the fact that I perceive different sorts of colours, smells, tastes, sounds, heat, hardness, etc., I rightly conclude that there are in the bodies from which all these diverse perceptions of the senses come, certain varieties corresponding to them, although perhaps these varieties are not in fact like them. ((My underlining; Descartes p. 159 (Sixth Mediation).))

(c) What is the relation between a person's mental events and the physical events which we may suppose occur in his nervous system, and by which the behaviour of that person might in principle be explained physically?

The central difficulty here is that there does not seem available for the dualist a clear sense in which mental and physical events could

interact. Moreover, it seems plausible and likely that all physical events in the body have physical causal antecedents and consequents: - in other words, it seems to be neither conceivable nor required for the explanation of physical events in terms of causal antecedents, that mental events link causally with physical events in a chain of events. This sort of consideration led of course to the development by Geulincx of the occasionalism latent in the Cartesian account; and, more generally, to psycho-physical parallelism.

(iv) There is the problem of deciding which physical structures sustain mental events in addition to physical events. Descartes wanted the dualism to apply only to human beings, and not also to animals. But it seems inconceivable that any physical differences between men and beasts (or any other physical structures) could ground the supposition that mental events occur in addition to physical events only in the former.

But despite all these difficulties attached to ontological dualism, it is rather difficult to abandon the view happily, chiefly for the reason that attempts to replace it by a monistic account do not appear altogether satisfactory: - it is rather difficult to eliminate or redirect the strong intuitive pressure leading to the supposition that there actually occur in people events such as the experiencing of pains, which are radically different from any physical events which could be conceived as occurring in the body. A physicalist monism would require that just physical events occur in people. But it is rather difficult to undermine the conviction that an event actually occurs somehow "associated" with a person's body when for example that person experiences an intense pain, which could not be any physical event observed to occur in the person's body.

1.2 Approaches to Physicalist Monism

The two most discussed approaches to physicalism are behaviourism in its various forms, and the mind-brain identity theory.

Logical behaviourism proposes that the meaning of all mental attributions can be analysed so as to show that they are assertions about behaviour. But logical behaviourism is most convincing where it is applied to mental attributions which cannot in any case be taken as asserting the occurrence of specific mental events: attributions such as 'A wants to dance', 'A believes that lions are harmless'. Behaviourism of the Rylean variety is notoriously unconvincing when dealing with the having of sensations and mental perceptual events.

In contrast the mind-brain identity theory, which was proposed in response to certain inadequacies of logical behaviourism, notably the implausibility of the claim of general meaning-reducibility, does not claim that expressions purporting to refer to mental events can be analysed so as to be equivalent in meaning to expressions which mention only physical entities. Rather, on the analogy of certain scientific hypotheses, 'A's visual imaging while he was seeing the water boil', for example, is taken to refer to a neurophysiological event whose description in neurophysiological terms is not yet known. But there are strong grounds for resistance to this sort of view.

One view which can be classified as a behaviourist view, and one of great subtlety and power, is Wittgenstein's - at least, a view which can be extracted from his later writings. On this view 'A's visual imaging' does not, and indeed no expression could, coherently refer to the sort of mental event envisaged as occurring by the dualist. As we shall see, though, Wittgenstein's view cannot readily be understood as claiming meaning-reducibility: expressions such as 'visual imaging' have not the same use as expressions which describe the behaviour of a

person. Nevertheless the coherent use of 'visual imaging' cannot, according to Wittgenstein, be that supposed and required by the dualist. Wittgenstein's arguments can be interpreted as exploiting the problems (i) and (ii) mentioned above as troubling the coherence of dualism.

In Part 2 I will be concerned with assessing the impact Wittgenstein's views have on ontological dualism. This discussion will concentrate on what I called the incommunicability component of privacy, for it is with respect to this aspect that we can most conveniently view Wittgenstein's arguments as suggesting serious difficulties regarding the coherence of ontological dualism. My procedure will be this:

(a) I will put in abeyance the problems (iii) and (iv) mentioned above, and try to project a view which appears at least to surmount problem (i) - an account in which (supposedly) communication is possible despite the condition that each person has a necessarily private view of the world mediated through his private sensations. In so doing I will argue that such an account requires that something about a person's private mental events must remain incommunicable by him to others.

(b) This will take us up to 2.4, where I will argue that if Wittgenstein's arguments can undermine the coherence of the sort of account presented so far, then not only is ontological dualism incoherent, but our concept of mind could not then obstruct physicalism in its weaker form. There I will isolate the supposed privacy of the mental as the crucial obstacle to physicalism. I will also suggest there that the weaker form of physicalism is sufficient for the stronger form (though I will not argue for this in any detail until Part 5).

(c) Then in 2.5 and 2.6 I will examine how Wittgenstein's arguments affect the dualist picture projected. My conclusion there will be that his arguments do succeed in rendering ontological dualism incoherent; but that his line of argument leaves an important gap regarding what is

to be the correct monistic account. The logical force of his argument will leave us with two options, both apparently counter-intuitive. One is to abandon the idea that we can be introspectively aware of and report the occurrence of mental events occurring in us (essentially a behaviourist line). The other is to suppose that such mental events would not be in principle private (essentially the line taken by the mind-brain identity theory, and the line that I will take).

I will only very briefly suggest the weakness of other behaviourist lines, partly because I consider Wittgenstein's as by far the most powerful of such views, and partly because there is a general body of opinion that approaches such as Ryle's do not adequately render the mental harmless to physicalism. Indeed, this inadequacy, as I have mentioned, is one of the main reasons why the mind-brain identity theory was developed. Wittgenstein's line, moreover, tends to be greatly underestimated as an approach towards physicalism, especially by identity theorists , whose discussions tend to be naive to the full impact of Wittgenstein's views.

In Part 3 I will discuss the mind-brain identity theory. I will find that the intuitive resistance to this sort of approach produces objections which do not seem to have been satisfactorily met by its proponents. And, for reasons I will give, although these objections can be interpreted as arguing for ontological dualism - a position rendered incoherent by Wittgenstein - Wittgenstein's arguments do not explicitly help us to disarm them.

But in Part 4 I will suggest an account by which a form of the identity theory can be supplemented so as to undermine both the intuitive resistance and the outstanding objections to it. This account, which I will call the augmented Identity Theory, is based on a re-examination of the realist distinction between perceptions and what is perceived. As

I have already hinted, ontological dualism is closely related to a certain problematic realist view. But we will find not only that such a realist view is problematic (being in conflict with the Wittgensteinian results we will have seen earlier), but also that the progression to ontological dualism from this problematic view is <u>in any case</u> muddled. I will try to give a more satisfactory realist account in a monistic setting.

In Part 5 I will discuss in more general terms the approach to physicalism made possible by exposing this muddle.

During the course of the thesis I will hint at a view for which I will argue in Part 5: that a <u>conceptual</u> dualism involving the contrast between scientific discourse and the peculiarities of mental discourse (such as intentionality) can be rendered quite harmless to physicalism in the strong form. There are various aspects of our "mental life" which I have not the space to touch on, and which should according to my account fail to present problems for physicalism (in the strong form). Perhaps the most important of these issues concerns the notion of freedom of will. I will very briefly mention this issue in my conclusion.

NOTES TO PART 1

[1] I will generally refer to dualism of mind and body simply as dualism.

[2] I will also discuss there in some detail other distinctions which have been used in the justification of dualism.

[3] Clearly we must take the dualist's 'perceive' (and its derivatives) in an extensional sense, rather than in the sense in which we can perceive that an event is occurring.

[4] Where it is obvious from the context that I'm talking about events which are in principle private to one person, I will often use the less cumbersome expressions 'private mental event' or 'private event'. And where it is clear that I'm talking of imperceptibility in principle, I will often just use 'imperceptibility'.

PART 2

PRIVACY LANGUAGE AND PHYSICALISM

The notion of privacy is perhaps most deeply discussed by Wittgenstein. His remarks suggest (a) that natural language does not contain references to private (mental) entities; and (b) that no reference is possible in any language, natural or "unnatural", to private entities.

In 2.4 I will argue that the latter view, (b), entails that there are not and cannot possibly be any private entities; and entails that an ontological dualism of the mental and the physical, in which the former is segregated on the grounds of its privacy, must collapse.

I will also argue in 2.4 that a claim maintaining ontological dualism of the mental and the physical must either be based explicitly on a segregation of the mental on the grounds of its privacy, or at least implicitly require that basis for segregation. (will also suggest in 2.4 that only ontological dualism can logically obstruct physicalism even in its stronger form; and that therefore if (b) is correct there should be no logical obstacle to physicalism.)

In 2.5 and 2.6 I will argue that Wittgenstein's arguments at best prove (b), and hence show that we cannot construe the relation between a person and his mental events as involving logically privileged access; and at worst render it very difficult to make sense of the claim that a person has logically privileged access to his mental events.

This result, coupled with the arguments of 2.4, suggests that arguments for ontological dualism contain or require an incoherent notion of mental events. But we will find that this consequence of Wittgenstein's line of argument leaves us with a difficult dilemma. On the one hand

Wittgenstein suggests that a certain notion of mental events as in principle private is incoherent. While on the other hand there is a very strong intuitive pressure which seems to require us to continue to maintain this problematic notion of mental events.

The view that mental events are in principle private is one which many philosophers have held with conviction. In order to give sympathy to this conviction, and because the argument for (b) is essentially a reductio taking this conviction as giving a premiss to be rendered incoherent, I will initially try to project the view that mental events are private in as favourable a manner as I find possible. I will project a view (which I call Stage I) in which private mental events are admitted and in which both (a) and (b) above are denied. This will inevitably involve the initial protection of this account from Wittgenstein's severest assault, the Private Language Argument.

A further justification I have for this prolonged reductio is this: one of the reasons why the bundle of arguments generally taken to constitute the Private Language Argument seems unclear, as regards both its potential consequences and whether it entails these consequences, is that not all the strands in this bundle are equally powerful - in particular, some strands in Wittgenstein's line of argument at least apparently do not achieve (b). But by means of the rather delayed reductio I give I hope to clarify some of the issues involved.

The discussion below can be managed with a fairly loose concept of privacy. Essentially (as indicated in Part 1) we will take a person's mental events to be private to that person in the sense that they are necessarily imperceptible to the senses of others, but their owner has logically privileged access to them. Moreover, it is supposed that in some sense a person's private mental events are incommunicable to others: this sense will tighten as the discussion proceeds.

2.1 Two Stages

Wittgenstein liberated the concept of meaning from the notion that there must be some "object" which is the meaning of a word or expression. Over-emphasis on referring expressions, in particular proper names, led Russell to the extreme of identifying meanings with referents. But the "object" view of meaning was present to some extent even in Frege's writing, despite his careful distinction between the sense and the referent of a referring expression. For Frege took the "sense" of an expression as an "object" in more than a superficial grammatical way. This is indicated both in Frege's illustrative analogy involving viewing the Moon through a telescope (cf. Frege (2), p. 60), where he compares the reference, sense and (private) "idea" to the Moon, the real image projected by the object glass in the telescope, and the retinal image of the observer, respectively; and in his remarks on the sense of indicative sentences (whose sense he calls "thought"), where he concludes that since senses are neither private "ideas" nor part of the world, they belong to some third realm (cf. Frege (3), p. 29). So though Frege gave meaning a non-psychologistic account, in that both a language user's subjective connotations and private mental events are irrelevant in fixing meanings, he did not go as far as Wittgenstein either in abandoning the search for an "object" which the meaning of an expression must be, or in banishing private mental events.[1]

For Wittgenstein of course the sense or meaning of a sentence or language move is likened to the function it has in the machinery of language. To know how sentences work as moves of language, and how words and expressions work or take part in language moves, is to know what they mean. And Wittgenstein (mainly in his later writings) was very much concerned with how the criteria for the correct use of expressions operate. He investigated how certain rule-giving contexts for

the application of certain concepts establish limits for the domains within which these concepts can sensibly be applied. So, by examining the legitimate contexts for the use of concepts expressed by 'know', 'mean', 'sensation', 'mind', 'rule', 'determine', 'proof', for instance, he was able to reconsider some traditional philosophical problems involving these concepts.

Of particular relevance to our discussion is the role, if any, of private (mental) entities in the workings of language. It is possible to abstract two stages in the diminution of the role of private entities in language; and it is clear that Wittgenstein held the view which I will refer to as Stage II, in which private entities disappear altogether. However, many of his remarks suggest no more than a Stage I picture, which roughly corresponds to the view of Frege's in which there are private entities, but they play a very restricted part in natural language.[2]

The way I educe these two stages entails that the first requires the occurrence of private mental events, while the second denies their occurrence. Stage I essentially allows reference of a sort to private events, but allows the latter no role in <u>fixing meanings</u> in natural language. Whereas the progression to Stage II, chiefly by means of the Private Language Argument, involves denying altogether the occurrence of private events. Much discussion on this topic fails to distinguish clearly between these two stages, and lumps together remarks of Wittgenstein's as entailing Stage II when in fact they can be divided in respect of the stage which they entail. I'm not suggesting that Wittgenstein himself was particularly conscious of or concerned with this division, since he was anxious to undermine <u>any</u> account admitting private events. Trying to apply this division to Wittgenstein is thus necessarily somewhat Procrustean - but it may help to clarify the issues.

Of course some philosophers, Locke being the most commonly cited example, have espoused a pre-Stage I view in which private entities play a dominant role in fixing meanings in natural language. We will discard this view almost immediately and without ceremony.

I will now proceed to an account of Stage I, one in which I will be relatively uninhibited by the possibility that Stage II will undermine it.

In the account below I will use 'private entities' to cover private mental events and types and properties of private mental events. And I will use 'private sensation of red' rather loosely, without committing myself as to whether or not it is events, such as the having of a private sensation, which are thereby referred to. My justification for this looseness is that any further precision is not relevant to the main issues covered in Part 2, and that I will often discuss views in which this looseness is also present, and with which it would be more cumbersome to engage if I were using 'private sensation' more precisely.

2.2 Reference to Private Entities: Stage I

Problems of reference are naturally a major preoccupation in semantics. But a large body of discussion in semantics is able to leave out consideration of private mental processes, and take as its starting point a "public" world with "public" objects.[3]

But in Stage I each person has a private view of the world, and we will have to dig a little deeper to understand how communication is possible. First, however, we must consider reference to private entities, and in order to outline how it is to be achieved we must say just a little about referring expressions.

In the case of proper names, which can conveniently be considered as semantically unstructured referring expressions, their meaning is

fixed by their referents: in the simple sense that criteria for the correct use of a proper name cannot be given independently of its referent. That is to say, we have to specify the referent, either ostensively or by description, in order to determine the correct use of the name.

In the case of semantically structured referring expressions the situation is rather different. Here the referent is prescribed by the structure of the expression. Thus in operating both what Donnellan calls the attributive function of a definite description, in for instance 'Smith's murderer is insane' ('Whoever murdered Smith is insane'), and the clear referential function of 'the book I see in front of me' (cf. Donnellan, pp. 285 - 286), any object which satisfies the identifying restriction prescribed by an expression can be correctly referred to by that expression. The referent of a structured referring expression does not fix the meaning of the expression: - for we may know the meanings of such expressions, how to refer with them, independently of knowing what are their referents; indeed we may know the meaning of such expressions, how to use them, even when there is no object which satisfies the conditions for being the referent of one such expression, (e.g., 'the tunnel under the English Channel').[4]

What we find in Stage I is that though private entities could fix the meaning of expressions in a private language, they do not and cannot fix the meanings of expressions in a natural language in which communication is possible: proper names whose meaning is fixed by their private referents could only be expressions in a private language. But we can nevertheless refer to private entities in natural language indirectly, by means of suitably structured expressions, the meanings of which are not and cannot be fixed by the private entities they refer to. So, for instance, 'my private sensation of red' will succeed, according to

Stage I, in referring to something private; but its meaning is not fixed by my private sensation of red.

2.21 Stage I starts off from this point, that there must be publicity of criteria in natural language.

Frege made this point in Frege (2) and (3).[5] In his earlier paper, however, Frege mingles the point that private entities (which I will henceforth refer to as P's) do not fix meanings in natural language with another point: he draws the contrast between the sense of a sign, and the ideas associated with it by different users: 'A painter, a horseman, and a zoologist will probably connect different ideas with the name "Bucephalus".' ((2), p. 59). In this example 'idea' is used so that we could allow the three users to communicate and further describe their associated ideas. In the next paragraph, however, Frege goes on to draw the distinction between the sense of a sign which can be grasped by any number of users, and the "idea", this time meant specifically as a P (a private entity), which is private to one person.[6] Frege emphasises the second point in Frege (3). He asks the question whether "thoughts" (intended to cover just the senses of assertive or indicative sentences, though in fact used more loosely) are "ideas" (as P's). Frege makes the following remarks in his ensuing argument against senses being ideas (P's), which can be interpreted as a denial that meaning is fixed by P's - that's to say, we may extract this point without going along with Frege's Platonic conclusion assigning meanings to some "third realm":

> I go for a walk with a companion. I see a green field, I have a visual impression of the green as well. I have it but I do not see it. (((3), p. 26)

> My companion and I are convinced that we both see the same field; but each of us has a particular sense-impression of green. I notice a strawberry among the green strawberry leaves. My companion does not notice it, he is colour-blind. The colour impression, which he receives from the strawberry, is not noticeably different from the one he receives from the leaf. Now does my companion see the green leaf as red, or does he see the red berry as green,

> or does he see both as of one colour with which I am not acquainted at all? These are unanswerable, indeed really nonsensical, questions. For when the word 'red' does not state a property of things but is supposed to characterize sense-impressions belonging to my consciousness, it is only applicable within the sphere of my consciousness. For it is impossible to compare my sense-impression with that of someone else. ((Ibid., p. 27.))
>
> I pick the strawberry, I hold it between my fingers. Now my companion sees it too, this very same strawberry; but each of us has his own idea. No other person has my idea but many people can see the same thing. ((Ibid., pp. 27 - 28.))
>
> Is that lime-tree my idea? By using the expression 'that lime-tree' in this question I have really already anticipated the answer, for with this expression I want to refer to what I see and to what other people can also look at and touch.... If my intention is realized when I refer to something with the expression 'that lime-tree' then the thought expressed in the sentence 'that lime-tree is my idea' must obviously be negated.... I have, naturally, got an idea then, but I am not referring to this with the words 'that lime-tree'. Now someone may really want to refer to one of his ideas with the words 'that lime-tree'. He would then be the bearer of that to which he wants to refer with those words, but then he would not see that lime-tree and no-one else would see it or be its bearer. ((Ibid., p. 28.))

Frege summarises: 'If someone takes thoughts to be ideas, what he then recognizes to be true is, in his own view, the content of his consciousness and does not properly concern other people at all.' (Ibid., p. 29.)

Implicit in the making of these sorts of points is the notion that P's at least seem to play a role in "generating" our behaviour (including linguistic behaviour). It would be maintained that when we react voluntarily to our environment we at least seem to do so on the basis of private events occurring in us. In other words, private events in a person seem to take on, at least partially, the role that, it may be hypothesised, neurophysiological events play. This naturally presents a problem - a mind-body problem. This view of the role of P's would of course be an expected consequence of psycho-physical parallelism.

We may put the mind-body problems concerned in abeyance for the time being. But we can suggest as a result a preliminary partial analogy both for the role of P's regarding the meaning of expressions in natural language, and for the mechanism by which they are specified or referred to.

Suppose a machine was programmed to respond to red and green stimuli (by means of a photoelectric device) with the respective outputs 'red' and 'green'. These outputs are generated on the basis of certain physical (electronic) events in the machine. But these physical events are not relevant to the meaning of the symbols constituting the outputs. For there are many different sorts of event which could generate these outputs in the same situations in which they are generated by the machine - and these outputs would thereby have the same meaning. For what is crucial to the meaning of 'red' and 'green' is the situations in which they are outputted, not the mechanism by which they are outputted. Furthermore, we may specify indirectly an electronic event in the machine as the event which takes place when the machine responds to, say, the red stimulus. Similarly, we could specify the neural event that occurs in me typically when I see red things: 'the neural event that occurs in me typically when I see red things' refers to something which, contingently, we have no access to. And its referent is quite unimportant in fixing the meaning of the expression used to specify it.

This is only a partial analogy, in that in principle we could specify the electronic or neural events directly by their intrinsic properties; whereas in principle no further description of my private sensation of red would be available.

We must now consider in more detail how P's can take no part in fixing meanings in natural language, and thus explain how communication would still be possible.

2.22 There is some ambivalence in the way such expressions as 'sensation of red', 'experience of pain' are used in ordinary language. Often they are used so that it is perfectly sensible to say, for instance, 'A and B both had the same sensation when they stopped looking at the light - they

both had the sensation of red'; or 'A and B had the same experience of pain in their chests'. But in other contexts it would be insisted that, for instance, A and B could not have the same sensation of red - it is not sensible to suppose they could. This is normally emphasised by saying that there can be no criteria by which we could judge that A and B have the same private sensation of red.

Now, partly to emphasise that a private sensation is specified, and partly to clarify the structure of expressions which are supposed to refer to P's, I will use the following conventions:

(a) 'AP(X)' abbreviates 'the type of private event occurring in A when X',

where 'X' is a specification of some publicly describable state of affairs.

(b) 'v Z' abbreviates 'he/one sees or has the illusion that he/one sees Z'.

(c) 'red', occurring inside the brackets of an expression of the form 'AP(X)' abbreviates 'a red object in standard light conditions'; 'yellow' etc. abbreviate correspondingly.

(d) I will often leave out the person-prefix when not required by the context, and use for instance 'my P(X)', 'the P(X)', leaving as understood that a P(X) is private to one person, and leaving as understood the appropriate grammatical changes required (such as omission of the definite article, change in pronoun, etc.) to interpret the abbreviations.

So, by replacing 'X' by 'v Z', and 'Z' by 'red', we form 'AP(v red)'. 'AP(v red)' is supposed to mean roughly the same as 'A's private sensation of red'. Although the correspondence is very rough, it is sufficiently close to bring out by means of the abbreviations the points I want to make.

The meaning of 'AP(v red)' is not private to A, even though its referent is. AP(v red) does not fix the meaning of 'AP(v red)' in much

the same way that Smith's murderer and Smith's girlfriend do not (respectively) fix the meaning of 'Smith's murderer' and 'Smith's girlfriend'. But in contrast to Smith's girlfriend whom we may publicly specify, as Charlotte for instance, if A specifies AP(v red) by an expression whose meaning <u>is</u> fixed by its private referent, then the expression will have a private meaning.

Suppose I use 'P'' to label my P(v red). We can exhibit the privacy of meaning of 'P'', in contrast to the publicity governing 'P(v red)' as follows. Suppose my P(v red) were to change according to the day of the week. Well, whichever day of the week it was, if I saw an object generally agreed to be red, a fire-engine say, I would still have a P(v red); since to use natural language correctly I would have to describe a fire-engine as red whatever day of the week it was. And if I said my P(v red) was P' on Monday, and P" on Tuesday, the criteria for using 'P'' and 'P"' would be private to me - their use is bound by their referents.

Essentially, the expression 'X' in 'P(X)', indexing or specifying a P, is bound by public criteria: is part of natural language for which the criteria of correct use are public. And though there can be public criteria for using the symbols in the construction 'P(v red)', there could not be for the use of the indices ''' and '"'. In contrast, supposing Smith to be rather friendly and of regular habits, which girl was Smith's girlfriend might also depend on the day of the week. But we have the facility in principle to communicate something in saying that Smith's girlfriend was Charlotte on Monday, Cherry on Tuesday, and so on.

So, what it is to be a P(v red), a private sensation of red, does not depend finally on any intrinsic private feature, whereas what it is to be P' does. Another way of exhibiting this is to consider the possible

case where from birth, instead of the P(v red) I do have, say, P', I had had P". Well, I would still have called it my P(v red), and it would have done just as well to underlie my judgement that fire-engines are red.[7]

Using examples of different P's in one person avoids the difficulty, in trying to make the same sort of point, of saying that AP(v red) might well be different from EP(v red). For it isn't at least obviously without sense to suppose that A can compare P' and P", whereas it is clearly not sensible to suppose that AP(v red) and EP(v red) can be compared. However, by extrapolating from a case where we can more readily understand "sameness" and "difference", where referents are only contingently private, the same point can be achieved - as in Wittgenstein's famous beetle-in-the-box analogy; though even here Wittgenstein is suggesting more than just a Stage I point:

> Suppose everyone had a box with something it it: we call it a "beetle". No one can look into anyone else's box, and everyone says he knows what a beetle is only by looking at his beetle. - Here it would be quite possible for everyone to have something different in his box. One might even imagine such a thing constantly changing. - But suppose the word "beetle" had a use in these people's language? - If so it would not be used as the name of a thing. The thing in the box has no place in the language game at all; not even as a <u>something</u>: for the box might be empty. - No, one can 'divide through' by the thing in the box; it cancels out, whatever it is.
>
> That is to say:if we construe the grammar of the expression of sensation on the model of "object and designation" the object drops out of consideration as irrelevant. ((I, 293.))[8]

The outside of the box, the "behaviour" of the box, can be described; but the beetle can only be individuated in terms of the box, and if the beetle had "states", these could only be individuated in terms of box "behaviour". Wittgenstein observes that we learn all our words and the concepts expressed through them in such a manner that, in the analogy, the outside of the boxes gives the final criteria governing the use of our words. But according to Wittgenstein the result of this is that

absolutely nothing can be said about what's inside the boxes, because we have learnt to use 'something', 'nothing', 'describe' so that their sensible use is confined to the world outside the boxes. Wittgenstein frequently insists on the restriction of the sensible use of terms and the concepts they may express to the domain for which we have criteria governing their application. Concepts whose governing criteria are given in language games set in the public world - concepts of natural language - cannot sensibly be made to apply in areas which never enter into these language games. So the insides of the boxes could not even be distinguished as containing a "something" rather than a "nothing" - the application of such a distinction would exceed the limits only within which we have secured its meaning. As I will suggest below, however, this extreme view is vulnerable to certain qualifications which will enable us to say something about P's.[9]

Before considering these qualifications we can briefly note that, in as much as we can extract a Stage I point from the beetle-in-the-box analogy, we can analogously use the expression 'p(box)' to specify the beetle: as P(X) stands to X, so p(box) (i.e. the beetle - whatever's inside the box) stands to the box - and in neither case can we go on to describe publicly the private entity in question.

2.23 So far I have presented a rather simplified picture of Stage I, which can be summarised just by the following points:

(a) Events private to each person do occur.

(b) These private events seem to provide the basis on which a person reacts to his environment.

(c) But these private events and their private features play no part in giving the criteria determining the meaning of the terms in the language with which people communicate; and they cannot be described in the way that the public world can be described.

As I mentioned earlier, we will leave alone for the moment the problem of (b). As regards (c), two important qualifications must be introduced, the second of which I will make in 2.24. The first involves the extent to which we can say anything about P's. We urgently need further comment on this, since it will not have gone unnoticed that already in setting up examples and so on, we need to say such things as 'P' is different from P"'. In other words, we have already presupposed that we can say <u>something</u> about P's. And if we can talk about P's at all we must, in view of the direction and pressure of Wittgenstein's remarks, discriminate between what can and what can't sensibly be said about P's.

We have seen, for instance, that one could not communicate what the intrinsic feature of P' is by saying that P' = P(v red). Although one could communicate what colours one saw, one could not communicate what "private shades" one has privileged access to. And yet it seems at least at first sight sensible and communicative to say that, for example, my P(v red) <u>changed</u> this morning from P' to P". Why is this? Why, although I could not further describe or specify my P(v red), does it appear that I could communicate <u>something</u> about my "private environment" by mentioning the <u>change</u>?

The following suggestion could be made to account for this - a suggestion which at any rate looks as though it could be used by someone advocating a Stage I position, and which for the moment we can allow to bolster Stage I.

Some concepts we have can be defined without reference to any <u>particular</u> set of things, or any <u>particular</u> thing. For instance, '$(p.(p \supset q)) \supset q$' can be said to define the concept expressed by '$\supset$' without requiring any knowledge about the actual statements which could replace 'p' and 'q'. Thus it could be maintained that we can understand <u>something</u>

about a statement of the form 'p⊃q', even though we don't know what the statements p and q mean. So we have no immediate reason to suppose that '⊃' could not be applied, with the same sense in which it is used in connection with statements we can understand, to cases where we can't know what p and q say - because they are statements in a private language.

In other words, if we can conceive of a private language at all, we can only conceive of it in a way in which we could sensibly apply the laws of propositional calculus to it. And the same consideration applies to predicate calculus and set theory. Any concept which does not require in its definition the specification of any particular value of a variable 'x', or any particular predicate or set, could apply with the same sense in the case where the value of 'x' is P', say.

For instance, identity can be defined without requiring the specification of x and y in 'x=y'. Thus 'J=K' can say something, even if we don't know what the expressions 'J' and 'K' refer to - in this case then, what they mean. 'J=K', can be taken as saying, for instance, that whatever is true of J is true of K. In as much as we may understand this statement, 'J=K', we can also understand 'P'=P"', though only I may have access to the criteria determining the truth of 'P'=P"'. But I can thus communicate something in saying that P'≠P", or that yesterday P(v red)=P', today P(v red)=P'. (Of course it may turn out that even I cannot have criteria for determining the truth or falsity of 'P'=P"' - if the Private Language Argument is right. But for the moment we may assume that I do have such criteria; and allow that we can apply '=' with the same sense to both public objects and P's. For, for instance, it is only because we understand how '=' is meant to be used in 'AP(v red)= BP(v red)' that we can say that there cannot in principle be any criteria to determine the truth or falsity of that statement.)

Any concept which has "empirical content", like the concept of

redness, acquires its meaning from objects in the public world. But although the concepts in logic and set theory have of course public meaning, we understand their public meaning so that they are in an important sense independent of public objects, and must be applicable to any coherently specified set. Provided we can coherently specify a set of events as being in principle private, we can then apply these concepts to it. We may say, perhaps, that such concepts belong to the structure of language games generally, and so even to private language games.[10]

Wittgenstein of course would dispute all this. He could be interpreted as maintaining (a) that because we learn 'is the same as' so that it has public sense, we cannot use even this concept in connection with P's. And (b) even if I tried to, I would find I had no criteria to determine the truth of, say, 'P'≠P"'. We will leave consideration of (b) till later. For the moment, regarding (a), we must insist on behalf of Stage I that though 'red' could not possibly mean the same in both 'P' is red' and 'the fire-engine is red', this does not extend in an obvious way to, say, '≠', which it seems possible to apply in the same sense in both 'my P(v red)≠ my P(v yellow)' and 'Smith's murderer≠ Smith's girlfriend'. Moreover, we are unable to set up or conceive the notion of response to private stimuli unless we can apply at least the concepts of sameness and difference to those private stimuli. Indeed, as we might expect, applying the concepts of sameness and difference to the private environment is part and parcel of a view in which we are supposed to make sense of a subject who has logically privileged access to his private environment: <u>if</u> we can make sense of such a private environment, we must allow sameness and difference to apply among the constituents of the private environment. When we come to consider arguments for (b), we will find that we cannot make sense of such an environment.

But for the moment we may assume that we can apply at least some

concepts expressed in natural language to P's. For instance, we may accept as intelligible in Stage I a statement to the effect that for a red/green colour-blind person, A, AP(v red) = AP(v green).

2.24 In order to clarify some of the issues involved so far, and to introduce a further supplement to the initial account of Stage I, we may investigate a model of Stage I based on contingent privacy.

Extrapolating from models based on contingent privacy is logically dangerous, as I will mention presently. But it is required at least for a preliminary understanding of Stage I; and it proves useful as a tangible focus for exhibiting difficulties in Stage I.

The way the model is set up requires:

(a) Two "people" A and B, each with their own "contingently private" environment.

(b) That we assume that A's and B's private environments are in a sense (contingently) imperceptible to one another. Essentially, relative to A's access to his environment, his access to B's environment is indirect and partial.

(In as much as we will be able to understand the working of the contingent privacy model, we have to suppose that we can sensibly distinguish between A and his private environment (similarly for B). One of the dangers threatening Stage I we will pin-point later as involved in the assumption that, in extrapolating from a case of contingent privacy to a case of logical privacy, we can continue to construe sensibly the notions of a subject and his environment.)

Imagine two people, A and B, playing games of chess. Each player has his own "board" and "pieces" with which to keep track of the game. But we stipulate that neither player can see the other player or his board and pieces; and that no interaction is possible between the players except in the communication of "moves". These are communicated by means of some

device, inaccessible to the players, which connects each thing serving as a piece on A's board with a corresponding "piece-thing" on B's board. And this connection is such that if a piece-thing moves on one board, the corresponding piece-thing moves on the other board."

Now, we may refer to the piece-things on the boards in the following way: 'Ap(black queen)', say, refers to the piece-thing on A's board which plays the role of the black queen. This is analogous to the use of 'AP(v red)'.

To illustrate this further we may augment the model in the following way. We may suppose the players also each have "letter-things", corresponding letter-things being interconnected as the chess piece-things are. Now, by "moving" the letter-things, A and B can "talk" to each other about their games, pieces, positions and so on. The players could agree in their use of 'black queen', etc., and generally succeed in communicating. But it is irrelevant to the meaning of, say, 'bishop', what Ap(bishop) and Bp(bishop) are "actually like".

Of course, the same consideration applies to the letter-things: we may refer to a sequence of moves of A's letter-things by using 'Ap('bishop')', for instance. Again, what Ap('b') or Ap('i') is actually like drops out of consideration. When A and B communicate they do not communicate what their piece-things and private environments are actually like, but structural features of their private environments, for which the actual content drops out of consideration. When A describes the position of his bishop, say, he does not describe his actual private environment: when A says 'my bishop is next to your queen', 'my bishop' refers to the same thing that 'your bishop' refers to in B's response: 'Yes, your bishop is next to my queen'.

Suppose it occurred to A and B to talk about their piece-things. They might even use expressions like 'my p(knight)'. But, assuming they

had never shared a common environment, they could not describe their piece-things to each other any further. Suppose part of Ap(knight) "falls off", and there is no corresponding change on B's board. Then A cannot communicate what has happened. It could not mean anything to B, other than that Ap(knight) had changed, if A said that Ap(knight) was p' but is now p".

On the other hand, some of the concepts A and B use to describe what they <u>can</u> describe may be applied with communicable sense to their piece-things. For if A and B understand 'the position has changed', they can understand 'my p(knight) has changed'. Though what the change is cannot be described. Really this boils down to the fact that we cannot conceive of A and B agreeing in their use of language, or playing their games unless they could extend the use of the concepts of "sameness" and "difference" from their application to positions to their application to piece-things.[12]

This brings us now to the second qualification we have to make of Stage I. If we examine the chess model just given we notice that A's and B's private environments could not be understood to be just anything. In order for A and B to play their games they must have <u>corresponding</u> piece-things; and, for instance, it must be the case that Ap(knight) $\neq$ Ap(bishop), otherwise his game would be impaired by a condition analogous to colour-blindness. In other words, for A and B to play their games, and to be able to communicate in any fashion at all, there must be an isomorphism between A's and B's private environments. <u>Within</u> the restriction that A's and B's boards and pieces are sufficiently isomorphic to each other, and to a chess board and pieces, their private environments can be quite varied.

So taking the analogy back to reflect on Stage I, we require A's and B's P's to be isomorphic if there is going to be behavioural parity

between them. Essentially we have a position in which we communicate what we can in virtue of there being a common structure to our P's; and supposedly what is incommunicable is the "content" - the actual private qualities of our P's.

The role of isomorphism in an account of the Stage I sort, involving P's, is of course well-known. Frege for instance says:

> What is purely intuitable is not communicable. To make this clear, let us suppose two rational beings such that projective properties and relations are all they can intuite - the lying of three points on a line, of four points on a plane, and so on; and let what the one intuites as a plane appear to the other as a point, and vice versa, so that what for the one is the line joining two points for the other is the line of intersection of two planes, and so on with the one intuition always dual to the other. In these circumstances they could understand one another quite well and would never realize the difference between their intuitions. since in projective geometry every proposition has its dual counterpart;... Over all geometrical theorems they would be in complete agreement, only interpreting the words differently in terms of their respective intuitions. With the word "point" for example, one would connect one intuition and the other another. We can therefore still say that this word has for them an objective meaning, provided only that by this meaning we do not understand any of the peculiarities of their respective intuitions.
> ((Frege (1), pp. 35 - 36.))

Although the restriction of isomorphism is understood to be relevant in this context, it is not examined deeply enough.[13]

In the chess model we've just considered, we can put the incommunicability of the actual qualities of, say, Ap(bishop) in this way.

I'll call a "board" plus "pieces", together with certain relations between the pieces and between the pieces and the board, such that using these relations the board and pieces are suitable for playing chess with, a chess system. There could be any number of chess systems which could equally well furnish A's and B's private environments. If A had had a different private chess system, this would not have affected his moves: the difference could not have an outlet in the common chess and language moves of A and B. Similarly if A's private chess system were to vary in any way (say, by Ap(bishop) changing), but remained a chess system,

although it would in principle be communicable that something had changed, this change could not be described.

What is incommunicable, then, is the extent to which the private chess systems are not isomorphic. Another way of putting this is to say that what is incommunicable is which private chess system A (or B) has. For the "content" that is not captured by the "structure" of a chess system is precisely that which may vary from chess system to chess system - it is given by the respects in which chess systems need not be isomorphic.

We can present a simplifying analogy rendering the relation between 'P(X)' and 'X', and between say, 'p(knight)' and 'knight'. Suppose we have a certain structure S, which can be specified in terms of m positions or places in that structure. Any system isomorphic to S we may call an S-system. Now, for any particular S-systems, say AS and BS, we may distinguish between Ap(n) and Bp(n), where Ap(n) is the "thing" which occupies the n-th place in AS; similarly Bp(n) occupies n-th place in BS. There will be certain restrictions given by the structure S governing the n-th place. In the structure of chess there are correspondingly restrictions which determine the "knight-place" - from which we get the concept of the knight. So on this analogy, as p(n) stands to n, so p(knight) stands to the knight, and P(v red) stands to "v red".

Now, there is an important difference between this sort of model and the situation as regards P's.

In the chess model, there may be a lot about, say, Ap(knight) which A cannot communicate to B. There may be a lot about A's and B's private systems which is incommunicable. We can conceive many different chess systems with the same relevant structure. Which private chess system A has is incommunicable.

But there does not seem to be anything about my P's which would

correspondingly defy communication in our natural language. As Wittgenstein says: 'Funny that in ordinary life we never feel that we have to resign ourselves to something by using ordinary language' (Wittgenstein (3), p. 241).

We can imagine how Ap(knight) might be different without affecting A's chess moves: this is to say, we can imagine an alternative isomorphic chess system. But I cannot conceive how, for instance, my P(v red) could be different, without affecting my "moves": my judgements about colour. I cannot imagine how my P(v red) could be at all different from what it is without either my being colour-blind - failing to agree in a colour discrimination - or in principle being able to be caught out by, for instance, the way I arrange a series of colours according to some ordering relation. For instance, if <u>just</u> my P(v daffodil yellow) and my P(v navy blue) were to "swap", I would then have to judge a daffodil to be darker in colour than a sailor's uniform.

Colours form a continuous series, and it seems that colour-P's must also form a continuous series. And the disparity between the chess model and colour-P's reduces to this: with one possible exception (which I mention shortly), AP(v C'), where C' is a precise colour shade, could not be any different without failing to preserve the isomorphism of the structure of A's colour-P's. In other words, we cannot conceive of alternative isomorphic private "colour systems". The possible exception to this would of course be a total inversion of colour-P's. But it is not at all easy to conceive this.[14]

Clearly the disparity between the chess model and colour-P's rests on the "tightness" or degree of isomorphism, which differs in the two cases. In the case of colour-P's there doesn't seem to be anything we can point to as being the "content" which is not captured by the structure.[15]

Suppose we each have private collections of objects to be counted; we can reveal the number of objects presented to us, but not what they are (they could be marbles or potatoes etc.). There would then be properties of these objects we couldn't reveal by announcing the number of objects presented: these properties would be "lying idle" - they would not engage in the counting games (just as properties of p(knight) do not engage in the chess game). But suppose our counting games were more sophisticated, so that for each dimension in which an object in one's collection could be ordered, we could uniquely assign a number both to that dimension and to the position the object had in that dimension. Then a matrix of numbers we announce (the public move) could uniquely specify one and only one object in our private environment. This of course assumes we all have the same number of dimensions and the same degree of discrimination in each dimension. But assuming this, what would there be incommunicable?

Again, in the chess model, A may alter Ap(knight) in various ways without there being a corresponding alteration to Bp(knight). But suppose that in whatever way A altered Ap(X), there was a corresponding alteration in Bp(X). This is to suppose a total isomorphism between A's and B's private environments. Then it seems there would be no "idle" and hence incommunicable features of A's or B's private environments. For the incommunicable features of, say, Ap(knight) were just those that were not captured by the isomorphism between A's and B's private chess systems.[16]

2.25 We have seen that a closer examination of a Stage 1 account adequate to represent how we communicate reveals rather troubling features.

First, by considering a simple model based on contingent privacy, we note that in order for some communication to occur there must be some

isomorphism between the private environments of the communicators. And the extent to which features of a private environment are not communicable is dependent on the common structure being sufficiently loose not to determine with precision what, say, Ap(knight) is actually like:- the extent to which the private environments are not isomorphic.

But secondly, we found that a total isomorphism is required between our colour-P's to account for the sort of behavioural parity obtaining in our judgements of colour; for the slightest change in my P(v red), say, is matched by a corresponding change in other normally sighted peoples' P(v red)'s. And this seems to apply to all our P's given the sort of parity obtaining in our judgements of size, weight, loudness, pitch etc. If I try and search for something about my P(v red) analogous to, say, the "height" of a piece-thing serving the function of a knight in chess, or the shape of a certain symbol in a certain code, I can't find it. But then if the isomorphism between our P's is so fine that every private quality is captured by the common language, what exactly is incommunicable?

Well, this is really not at all bad for Stage I. Because one of the tensions in a position admitting the occurrence of private events is how to reconcile this with the fact that we seem to be able to communicate without the sort of hindrance privacy might be supposed to create. We can make this conciliation in the following way.

The crucial factor which removes the hindrance in communication is the empirical fact that, among normal people, potentially any change in, say, A's P's can be matched by a corresponding change in B's P's. Nevertheless there remains a residual incommunicability: remembering that this incommunicability is given by the possibility of an alternative isomorphic "private system", the incommunicability now hinges on there being another totally isomorphic "private system" - such as a colour-P

inversion. In which case which of the two inversions obtains for A could not be communicated.

I will consider in the next section a much more plausible case of P-inversion.

2.3 Residual Privacy and the Ozma Problem

Assuming a perfect colour inversion is possible, my P(v red), for instance, and generally any colour-P, could be one of two P's having the corresponding position in the two isomorphic colour-P spectra. Either of these two inversions could ground behavioural parity between me and normally sighted people. This possibility would not be allowed as coherent by Stage II.

The different case which Frege considers, involving A and B with "different" but isomorphic private systems (cf. 2.24), is considered by Dummett ((2), pp. 638 - 639). Of course in this interpersonal case we have the difficulty that there can be no criterion for comparing A's and B's P's. But Dummett notes that implicit in Frege's example is the notion that A and B have private experiences, and that this is ruled out by Wittgenstein's comments.

In the intrapersonal case, Stage I maintains that A can compare his own P's, so A would be able to recognise the difference if there was a change from one P to another. The conception of such a change is again supposedly rendered incoherent in Wittgenstein's Stage II position.

I will now consider a case where there seems to be available a coherent conception of two isomorphic private "systems"; where which of these two I have is incommunicable; where it appears that I should be able to distinguish between the two, but I could only make that distinction privately.

2.31 Suppose that Y is something I see, a visual scene. I can describe

this scene in natural language: a banana is to my right of an apple. Now, I can imagine a scene Y' which is the lateral inverson of Y, and in which the banana is to my left of the apple - with the banana, the apple, and everything I see laterally inverted. We would of course be able to ascertain which of these two isomorphic configurations of objects I was seeing. There is no difficulty in accepting the distinction between Y and Y' as coherent.

But in Stage I we have to distinguish between Y and the private visual image I have of Y - P(v Y).[17] And similarly we would have to distinguish between Y' and my P(v Y'). And since Y and Y' are perfectly isomorphic, my P(v Y) and my P(v Y') must be perfectly isomorphic: otherwise I could not, from the Stage I account, judge Y and Y' to be perfectly isomorphic. Also in Stage I, I can make the private distinction between my P(v Y) and my P(v Y'). Suppose my P(v Y)=P^S, and my P(v Y')=P^T; then if tomorrow my P(v Y) changed from P^S to P^T, and for every Z my P(v Z) "swapped" with my P(v Z'), (where Z' is the lateral inversion of Z), then in principle, eventually (when my kinaesthetic P's had adjusted), I should achieve complete behavioural parity with other people - my movements, and my judgements of position would be normal. But, more strikingly, if from birth for every Z I saw I had had what we in fact refer to as P(v Z') instead of P(v Z), I would have behaved <u>exactly</u> as I have in fact behaved. So in the hypothetical case when I saw Y I would have had P^T instead of P^S, but of course I would still have judged the banana to be to my right of the apple.

Suppose the (private) relation R obtains between elements of my P(v Y), and the converse relation L obtains between the corresponding elements of my P(v Y'), such that as a result of how I've learnt to use 'right' I say that I see a banana to my right of an apple when I see Y. Then in the hypothetical case where from birth I had had P(v Z') in those

cases where in fact I have P(v Z), L would have obtained between my private visual images of the banana and the apple, but I would have learnt 'right' so that I would still have judged the banana to be to my right of the apple when I saw Y. The relations R and L are privately applicable by me, whereas those expressed by 'to the right of' and 'to the left of' are publicly applicable.

In Stage I the possibility of my applying R and L, and of distinguishing between P^S and P^T, are coherent. But we would have to accept that the use of R and L is private to me, and that the distinction between P^S and P^T is private and incommunicable. There is no way I could communicate to another person which was which, though I could communicate that they are isomorphic. On the other hand I could of course communicate whether I see Y or Y', and I could communicate everything there was to communicate about Y and Y'.

And suppose we induced the inversion of P's in A, either by means of inverting spectacles, or as in conceivable (on the hypothesis of psycho-physical parallelism), neurally. We could then distinguish between AP(v Z) without the inverting spectacles, or without neural tampering, and AP(v Z) with the spectacles or with neural tampering. But in Stage I there would be required a distinction private to A, and independent of whether he is wearing spectacles or not. That's to say, we could of course publicly distinguish between AP(v Z) and AP(v Z seen through inverting spectacles). But we could not tell whether the change from AP(v Z) to AP(v Z seen through inverting spectacles) is a change from AP^S to AP^T, or from AP^T to AP^S.

2.32 This restriction in Stage I on communicability associated with the private analogues "R" and "L" of "to the right of" and "to the left of", is very like what may be considered as a transposition of this incommuni-

cability:- the incommunicability associated with what is known as the Ozma problem. Gardner puts this problem in the following way:

> Is there any way to communicate the meaning of 'left' by a language transmitted in the form of pulsating signals? By the terms of the problem, we may say anything we please to our listeners, ask them to perform any experiment whatever, with one proviso: There can be no asymmetric object or structure that we and they can observe in common. ((Gardner, p. 176.))

Originally this problem arose in connection with the possibility of intergalactic communication and the left/right parity of physical laws. If no physical laws are as a matter of fact left/right asymmetric, how could we communicate the meaning of 'left' and 'right' to a group of beings with whom we have no perceived object in common by which to standardise our use of 'right' and 'left'? What we could do in this case is agree with these beings to use 'Right' and 'Left' by convention so that for instance one's heart is normally on one's Left (supposing these beings were appropriately similar anatomically), or that the north pole of a compass needle is deflected to the Left by applying current in a certain direction to a wire near it. Then we know that when these beings use 'Left' they mean by it the same as either our 'left' or our 'right', but we don't know which.[18]

This setting is analogous to Stage I, in that in Stage I A and B have no private events in common. And as 'to the right of' and 'to the left of' stand in relation to 'R' and 'L' in Stage I, so 'to the Right of' and 'to the Left of' stand in relation to 'to the right of' and 'to the left of' in the Ozma setting.

I will call this residual incommunicability inherent in Stage I, Residual Privacy. It boils down to the supposition within Stage I that which of two totally isomorphic visual imagings is occurring in a person could not be communicated or made manifest to someone else.[19]

As we see, the element of incommunicability which attends the notion that mental events are in principle imperceptible is much less serious

than suggested by primitive models of Stage I, such as the chess model - essentially because the isomorphism which would be required to be compatible with the parity in our behaviour would be much tighter than is required in the chess model.

Nevertheless there <u>is</u> an incommunicability component - rendered by Residual Privacy - involved in the claim that a person's mental events are in principle imperceptible to others; and this Residual Privacy is in direct opposition to Wittgenstein's Private Language Argument: for Residual Privacy requires the possibility of a private distinction by a subject between two isomorphic private visual imagings. It should be emphasised then that any view which claims that a person's visual imagings are in principle imperceptible to others requires the possibility that the person is able to make a distinction which is in principle private to him.

Thus, suppose it is maintained that the event a person, A, may be introspectively aware of as occurring in him typically when he sees Y is additional to the neural event typically occurring in him when he sees Y (supposing there is such a neural event). Let's specify that neural event as AN(v Y), in a manner similar to the specification of the private event. (I will later mention again this form of specification for neural events.) Then although we may be able in principle to perceive or at least to determine uniquely AN(v Y) as, say, N^R, we could not tell which of AP^S or AP^T is AP(v Y). But A could introspectively distinguish which of the two isomorphic visual imagings was occurring in him.

The following analogy might be helpful here. In addition to a photograph there is the actual scene which was photographed - this is to be analogous to the supposition that <u>in addition</u> to any neural events detected, there are mental events. Now suppose that (contingently) we could not compare the photographs with the actual scenes they depict -

all we could tell about these scenes was what we could tell from the photographs. This is to be analogous to the imperceptibility of the mental events (and also the Ozma separation mentioned just now). So in this case all we could determine was that the scene photographed was either "like" the photograph, or "the other way round". So we couldn't from a photograph determine which of two isomorphic scenes was depicted - though in principle we could determine everything else (visual) about the scene. And this is analogous to the Residual Privacy that must attend an account of necessarily imperceptible mental events. Pursuing the analogy, we could suppose there was a person who - having privileged access to the scenes photographed - could determine the "polarity" of the scene photographed.[20]

2.4 Stage I, Stage II, and Physicalism

There is in Stage I a distinction between two semantic conditions that can govern the use of expressions. I will call P-primacy the condition where the final criterion for the meaning of a term rests with a P, and is private to one person. And I will call B-primacy the condition where the criterion governing the use of a term is given by structural relations on P's. Because these structural relations can be common to various users, terms governed by B-primacy can have public meaning. In terms of the chess model, which though simplifying, illustrates the distinction involved, the analogue of B-primacy governs the use of 'knight', and the analogue of P-primacy governs the use of 'p'', where 'p'' refers to a piece-thing. We may say that the use of 'knight' is bound by a structural feature common to A's and B's private chess systems; and 'knight' may be used to specify a piece-thing: p(knight) is the piece-thing that takes the "knight-place" in the structure of the private chess system. The concept of the knight in chess is very

much like the concept of a number: in systems which have a common structure given by m places, 'the n-th place' corresponds to 'the knight-place', 'n' corresponds to 'knight', and 'the n-th thing' corresponds to 'p(knight)'. In all systems which have m places, the concepts of the n-th place, n, and the n-th thing are common to all. Whereas concepts used to describe the individual systems, by describing the individual n-th things, are special to each system.

And in virtue of being governed by B-primacy, 'P(v red)', 'v red', more generally 'P(X)' and 'X', are significant in natural language; whereas 'P'', which is bound by P-primacy, is not: its significance is confined to one person. The requirement of public objects or states of affairs by which to specify private sensations (sometimes called "raw feels"), is frequently noted.[21] In the case of other sorts of private mental events, such as trains of thoughts, the formulation of beliefs, desires and so on, part of the public specifying tag is provided by Brentano's intentional objects.[22]

So Stage I, essentially a development of Frege's picture, requires that P-primacy is possible; but that B-primacy operates in natural language. And we can use expressions governed by B-primacy, which have public sense, to refer to P's. This possibility of reference to private items using expressions which have public sense is sometimes overlooked, or criticised in too short a manner. Hopkins, discussing the possibility of public sense and private reference, says:

> ... for two persons to communicate (mean the same) by use of a word 'w' it must be possible for them to agree in judgments about w's, or at least judgments logically connected with these. So the referents of 'w' must be objects of public judgment. It would follow that persons could not communicate by words which referred to private objects, or to items independent of public circumstance. So the idea that sensations are independent or private would be inconsistent with the idea that we attach a common sense to sensation words. ((Hopkins, p. 131.))

P-primacy, which is roughly what Hopkins calls the recognitional

conception of meaning (cf. Hopkins, p. 128), certainly governs the use of 'P'', 'P"': these do not have public meaning in 'my private sensation of red is now P"; yesterday it was P''. But as we saw, the use 'my private sensation of red', 'my P(v red)', can be governed by B-primacy: can have public sense even though these refer to something private. Now as we shall see below, it cannot follow in this case that A can tell if it's <u>true</u> that B has a P(v red), a private sensation of red, for the simple reason that A cannot tell if B has a private sensation at all. But we can only say this if <u>we understand the meaning of</u> 'B's private sensation of red':- if we understand that it refers to something such that only B could tell if it occurs, but that, <u>assuming</u> it occurs, not <u>only</u> B can tell it's a private sensation of <u>red</u>. It is part of the public meaning of 'P(v red)' that it refers to something whose occurrence is unverifiable except by one person. By A and B agreeing on the use of 'AP(v red)' they must thereby agree that the truth conditions of 'AP(v red) occurs' are asymmetrically ascertained. But on the basic <u>assumption</u> that A has P's, then A and B can have equally good evidence for the occurrence of AP(v red). And it must therefore be maintained in Stage I that it is part of the public meaning of 'AP(v red)' that when we say that 'AP(v red) occurs' <u>we</u> must assume that A has P's.

We can further illustrate this aspect of Stage I by discussing some relevant lines of argument in Wittgenstein.

<u>2.41</u> Some of Wittgenstein's remarks which suggest or can be taken to suggest that there cannot be the privacy required by Stage I, are based not on the correctly engaging Private Language Argument, but on relatively much weaker reasons. These reasons tend to reduce to just the fact that natural language is public.

For instance, it might be taken as a criticism of the privacy of experience, that if only I know the meaning of 'red', we could not

communicate by means of 'red'; yet we undoubtedly do communicate by means of 'red'. Thus:

> 'Only you can know what colour you see'. But if it is true that only you can know, you can't even impart this knowledge nor can you express it. ((Wittgenstein (3), p. 270.))

But Stage I would not of course entail that only you could know what colour you saw. What Stage I entails is that only you can know what your P(v red) is privately like, only you can know if your P(v red)=P', say.

> ... compare secrecy with the 'privateness' of personal experience! In what sense is a thought of mine secret? If I think aloud it can be heard. ((Wittgenstein (3), p. 270.))

And I quote again:

> Funny that in ordinary life we never feel that we have to resign ourselves to something by using ordinary language (((3), p. 241.))

But as we have seen, the residual incommunicability present in Stage I would be quite compatible with this. Indeed, for most non-philosophical transactions it could be forgotten about altogether.

Many of Wittgenstein's remarks emphasise the requirement of public sense without, by themselves, requiring more than that P's could not enter into the criteria of public meaning:- without requiring Stage II, in which P's don't enter into anything, rather than Stage I. Elsewhere, though, Wittgenstein leaves Stage I behind.

> In what sense are my sensations <u>private</u>? - Well, only I can know whether I am really in pain; another person can only surmise it. - In one way this is wrong, and in another nonsense. ((I, 246.))

Here we must separate between Stage I and Stage II. For in Stage II it would be maintained with Wittgenstein that in the sense of 'pain' and 'know' that we have all learnt to use and understand, the interlocutor's remark is simply false: <u>public</u> criteria are just those which do serve to <u>establish</u> whether or not someone <u>is in pain</u>; so it is simply false that only I can know whether I am really in pain, since the criteria for establishing whether I am in pain are public. And in the way the

interlocutor's remark is intended to be about P's, it is nonsense, for then, according to Wittgenstein, it could not have public sense.

In Stage I the remark of Wittgenstein's interlocutor would be infelicitous, but for a different reason - the reason it would be infelicitous to say 'only I can tell if I have a private sensation of red'. Certainly, only I could tell that I have a P', or only I could tell that my P(v red)=P'. But we must not overlook this feature of Stage I, that anyone could tell, on the assumption that I have private sensations, that I have a private sensation of red, or a private sensation of pain, for these are public specifications of private entities. Certainly in Stage I we would hesitate to say that R has a private sensation of pain, if R, though behaviourally just like we are, was somehow known not to have private sensations - supposing R to be a robot, perhaps, in which only physical events occur. But on the supposition that A has private sensations, it would be quite misleading to say that only A could tell that he has a private sensation of red; for on the assumption that A has private sensations, only public criteria can warrant the statement that he has a private sensation of red.

Wittgenstein essentially has two routes to Stage II; one of them, the Private Language Argument, engages well with the issues involved, and if right certainly leads to Stage II. But the other route, which I have already hinted at, only leads to Stage II on the mistaken notion that we cannot have public sense with private reference. I will sketch a Wittgensteinian line of argument invoking this assumption, and then note how Stage I diverges from it.

In expressing the doubt as to whether other people have minds, we try to refer to something with 'mind' which is disallowed by the rules for its sensible use - or as Wittgenstein would say, the "grammar" of 'mind'. And this is because the criteria for possessing a mind are

given by public behavioural considerations; to atribute a mind to someone is in effect a move sanctioned by the display of "mind-possessing" behaviour, just as having an itch is determined in connection with the tendency to scratch. A person can hide the fact that he has an itch (and maybe could hide the fact that he has a mind), but then what he is hiding is not something which remains hidden when he scratches, or sincerely ('sincere' having its own public criteria for attribution) reports it. So that if someone showed our normal range of behaviour, then ipso facto he would have a mind. And of course to say that we can be deceived (by, say, a rather good dummy containing a tape recorder) is only to say that we can be deceived by behaviour which doesn't turn out to be of the appropriate sort. But we couldn't say that the behaviour was of the right sort, typical "mind-possessing" behaviour, and the subject still didn't have a mind, nor could we felicitously add 'and he did have a mind'.[23]

So we end up with a view here that there is nothing in the meaning of 'mind' which suggests that to have a mind one must have private goings-on. For we simply do correctly say that people have minds without requiring as evidence for this the occurrence in them of private events - how could these enter into the evidence? So on the supposition that it's sensible even to consider and formulate the case where A behaves just like someone with a mind, but there aren't any private events going on in him, we would still have to say that he has a mind, has sensations, is conscious, and so on.[24] What Wittgenstein would want to say is that because public criteria provide the grounds for determining whether or not I have a private sensation of red, the supposedly private referent we thereby try to mention is not really mentioned at all: I could very well "not have anything at all" privately going on in me, and it still would be correctly said that I have a private sensation

of red. That's to say, if the truth of 'I have a private sensation of red' is really determined by public criteria, we cannot use that form of words to say what we've been wanting to say in Stage I: for it would be judged to be a true statement even if, in the sense we've been wanting for Stage I, I don't have anything going on in me private at all - I just "behave". Whether I have in the intended Stage I sense something private or nothing private it would still be true, if there remains a use for 'private sensation', just in virtue of certain public circumstances, such as when I normally look at something red, that I have a private sensation of red. Nothing would be relevant to the truth even of 'I have a private sensation', other than what is public.

Now, in Stage I we have to disentangle the publicity of meaning of 'I have a private sensation' from the publicity with which Wittgenstein wants also to determine its truth. In Stage I both 'A weighs ten stone' and 'A has a private sensation of red' have public meaning. But in contrast to the former ascription, the public meaning of the latter entails that when we ascribe it to A we have to assume A has private sensations. And of course, implicit in attaching sense to this assumption is the requirement that we have to attach sense to the possibility that the assumption is wrong - even though everything we can take as evidence from the observation of other people is compatible with both the truth and the falsehood of that assumption. Stage I is very odd in this and other respects: - it is a position which it would be very desirable not to have to hold. But we must remember that while we are each individually unable to conceive how there are no events occurring in us which are imperceptible to others, and while we are not convinced that the occurrence in us of such events is not proven to be logically impossible, we seem to be forced to remain in Stage I.

So essentially Stage I survives the consequences of the line of

argument sketched against "other minds" scepticism, by insisting that there are some things we say, such as that A has private sensations, that A has a mind (in the sense of a private stream of consciousness), which <u>in virtue of their public meaning</u> are assumptions: the meaning, but not the truth of 'A has a private sensation of red' is fully determined by public criteria.

So in Stage I, despite the operation of B-primacy, despite the publicity of meaning in natural language, it remains possible to formulate the "other minds" scepticism.[25]

So in contrast to Strawson, who can be interpreted as taking the line that just the publicity of criteria operating in natural language vitiates "other minds" scepticism (cf. Strawson (1) pp. 48 - 49, (2), pp. 87 - 116), we have here that precisely because 'P(v red)' means what it does mean, and means the same in 'A has a P(v red)' and 'I have a P(v red)', we render "other minds" scepticism possible. For the truth of 'A has a P(v red)' entails the compound (a) A has a P, and (b) A's P is a P(v red). That (a) is true can only be an assumption on the part of B and others; but on the assumption that (a) is true, then A and B both have the same public evidence for saying that (b) is true.[26]

<u>2.42</u> A recurring problem for Wittgenstein is the difficulty of exposing the limitations of certain concepts whose application is delicately bound by their "grammar", when such exposition requires the use of those very concepts, often in a way which has to stretch them beyond their proper use. He is unable to formulate properly the mistakes he wishes to expose without making them himself. Thus he very often has two ways of talking, in one of which he uses concepts such as are expressed by 'know', 'mean', 'state of mind', 'sensation', 'determine' in their proper way, and in the other of which he is forced to use these concepts so that their use transgresses into misuse.

This is particularly striking in the case of 'sensation' and 'private sensation'. I will lead up to this with other examples of this difficulty in Wittgenstein.

For instance, when investigating the notions of giving and obeying rules, Wittgenstein writes:

> "But I already knew, at the time when I gave the order, that he ought to write 1002 after 1000." - Certainly; and you can also say you meant it then; only you should not let yourself be misled by the grammar of the words "know" and "mean". ((I, 187.))
>
> It may now be said: "The way the formula is meant determines which steps are to be taken". ((I, 190.))

Here he is reminding us of the normal (and correct) use of the concepts involved. 'Do I know that is a tree?' asked under perfectly normal conditions when a tree is in front of one, would contain a mistake - just as 'But <u>are</u> the steps then <u>not</u> determined by the algebraic formula?' (I, 189) contains a mistake. The mistake is of course that 'know', 'determine' are used in such a way that these are precisely the circumstances where it is correct to say 'I know that is a tree', 'the formula determines the steps'.

But of course Wittgenstein also finds a deep sense in which a formula <u>doesn't</u> "determine" which steps are to be taken, in which you couldn't "mean" it then. Even so, Wittgenstein is careful about this role for these concepts. Thus he says:

> It would <u>almost</u> be more correct to say, not that an intuition was needed at every stage, but that a new decision was needed at every stage. ((I, 186; my underlining.))

The reason why it would not be any more correct to say this than to say:

> "What you are saying, then, comes to this: a new insight - intuition - is needed at every step to carry out the order '+n' correctly." ((I, 186.))

is, as Wittgenstein knows very well, that it is <u>not</u> what we would call forming a <u>new</u> concept, or changing the concept, to follow the series

'2, 4, 6, 8, ..' with '10'. Rather, it is precisely what we would call forming the concept, following the rule, that we go on adding 2 (cf. also Wittgenstein (1) p. 143). Wittgenstein is not saying that we are making a new decision at each step; rather, that following a rule without making new decisions is not quite what it appears to be; and this can be illuminated by trying to use the concepts of 'decision', 'determine' at an unnatural level.[27]

The difficulty for Wittgenstein in the case of 'private sensation' is this: we do have a use for 'private sensation', but we cannot properly use it to refer to the sort of entity Stage I wants it to refer to, since for Wittgenstein we cannot in natural language, with public sense, refer to or describe in any way something which cannot enter into public meaning-criteria. So supposing Wittgenstein were to say, to make his point: even if no "private events" occur in A, but A behaves just as we do, it must still be true on certain occasions (when A sees a red object), that A has a private sensation of red - then according to Wittgenstein the sense of 'private event' occurring in the double quotes just now is not coherent if it is the sense required for him to make his point. And the tortuousness of his remarks on this topic are in a large part due to this fact, that he cannot in a coherent manner use natural language to express the view of privacy he is trying to combat. But of course he does, as indeed he has to, indulge in this incoherent use in order to engage with the position he attacks. This improper use occurs in the following passages, for instance:

> The essential thing about private experience is really not that each person possesses his own exemplar, but that nobody knows whether other people also have <u>this</u> or something else. The assumption would thus be possible - though unverifiable - that one section of mankind had one sensation of red and another section another. ((I, 272.))

> What am I to say about the word "red"? - that it means something 'confronting us all' and that everyone should really have another word, besides this one, to mean his <u>own</u> sensation of red? Or is it like this: the word "red" means something known to everyone; and in addition, for each person, it means something known only to him? (Or perhaps rather: it <u>refers</u> to something known only to him.) ((I, 273.))

But in the following excerpt he suggests how we cannot use 'sensation', indeed any meaningful word, in the way required to talk about something, in the deep (improper) sense, "private"; ('S' is supposed to be a sign for a private sensation):

> What reason have we for calling "S" the sign for a <u>sensation</u>? For "sensation" is a word of our common language, not of one intelligible to me alone. So the use of this word stands in need of a justification which everybody understands. - And it would not help either to say that it need not be a <u>sensation</u>; that when he writes "S", he has <u>something</u> - and that is all that can be said. "Has" and "something" also belong to our common language. ((I, 261.))

'Sensation' has public sense, so it cannot according to Wittgenstein refer to something "private". And if we try to use '<u>something</u> private', for that purpose, we still cannot succeed, for 'something','private' also have public sense. Thus Wittgenstein's interlocutor in many of the relevant remarks is like a man in a bubble who insists that he can come out of the bubble - but every time he thinks he's stuck a hand out Wittgenstein tries to show that the bubble film is still fitting his hand like a glove. The bubble distorts but doesn't break: - however we try and distort public language to refer to <u>that</u> sort of entity, the film of public sense prevents our success; the film clings as much to 'something', 'private' as to any other word with public sense.

And of course, Wittgenstein is also in the bubble, which makes his task difficult - he continues the last quoted passage (I, 261):

> - So in the end when one is doing philosophy one gets to the point where one would just like to emit an inarticulate sound. - But such a sound is an <u>expression</u> only as it occurs in a particular language-game, which should now be described. ((My underlining.))

This is a rather beautiful predicament, but not one we have to be in, unless the Private Language Argument is right. For as we saw, Stage I rejects the thesis that every mode of reference to P's is publicly senseless. Although 'P'' does not have public sense, as I've used it, 'my private sensation of red' does, and nevertheless succeeds in referring to the sort of entity which no one else has access to.

Before discussing Wittgenstein's other route to Stage II, the Private Language Argument, I will try and indicate some general features of Stage II.

2.43 What picture fits Stage II?

The Private Language Argument (which I will refer to as the PLA) essentially argues that it is impossible to use terms bound by P-primacy, because it would be logically impossible to have criteria of meaning under this condition.

If the observance of P-primacy involves a logical absurdity, there can be neither private languages nor private entities. For the existence of P's, like the existence of anything, requires that there be criteria for distinguishing one P from another - requires that P's should have properties. In the cases where we say of some sort of thing that it is not directly observable, we must nevertheless be able to set up criteria whereby the unobservable properties of these things are related to observable ones. Although, for instance, historical events are not directly observable by us now, there is a rationale by which such events are systematically related to events which we say are in comparison directly observable. But P's would be denied any properties at all, since there would be no criteria available to anyone for a P having one property rather than another; there would be no more sense to saying that a P has a certain property, or that a certain private event occurs in me, than to saying that something which was by definition undetectable

in any way was in point of fact a pink elephant. (The analogy is not so much in respect of the unobservability of the two, a private event and the pink elephant - for the private event would in a sense be accessible to one person - but in respect of the absence of criteria.)

So while Stage I has P's as semantically inactive in natural language, despite their existence, Stage II has "their" inactivity as due to "their" non-existence. And if Stage II is correct, then Stage I is undermined at a stroke: private events could not then just have a subservient semantic role - they would vanish from our account of people and their behaviour.

Stage II is really quite simple: there is nothing to people but their bodies and bodily states and behaviour. Or rather - for this is infelicitous since in natural language we have a use for sentences like 'A had a pain accompanying his behaviour' - in a deep sense there is nothing to people but what is physically accessible. Naturally Wittgenstein was troubled by how to depict this view, how to say what there isn't. But we can understand what he would want to say. His is an extremely subtle form of behaviourism (I hesitate to describe it as logical behaviourism since this is sometimes taken to imply a degree of meaning-reducibility) which we can illustrate by means of an example. The example is one which Wittgenstein might not find altogether attractive, but it is the only one I can think of that may help to understand his final position. At any rate it exhibits what I describe as Stage II. The example is one of a family of related examples which have frequently been used to make a family of points.[28]

Let's suppose that a certain organism is made just out of matter - the only events which occur in it are physical, and in principle perceptible; there are no private events occurring in it.[29] We could perhaps say that it is a sort of automaton . We may also suppose, for

simplicity, that this organism looks and behaves just like we do. For instance, it might say in certain situations: 'I'm in pain', 'I wish you could feel my toothache, it's dreadful', 'what a pity you can't see my after-image, it's such a nice colour', 'of course there's something over and above my behaviour'. All this is logically possible. If we knew that this was just a material thing with no private events occurring in it, only physical events, we might not know quite how to react. But suppose we didn't know this; then absolutely anything we could say of ourselves we could say of it: - we could discuss its feelings, problems, fears, desires and so on.

We can suppose further that there is a community of such organisms, all of which behave rather like we do. They might even discuss the possibility of a private language, whether each had its own private events. We don't need to ask why such organisms should behave in this way if they don't really have private events occurring in them: we must accept the logical possibility that they just do. They use 'private sensation', 'mind', and so on, in a perfectly sensible manner, just as we do. 'But what would such expressions used by these organisms mean?' Perhaps they mean what we mean by them - their use does not conflict with our use for them. 'But we mean "private events" really to refer to private events which occur in us, but do not occur in them'. This may be so. But the point is this: there don't have to be any private events for the words we use to be used as they are used. Absolutely all the words we use to describe ourselves, our minds, and so forth, could be used in exactly the same observable situations we use them in, without there being anything but physical events.

So if we step back a little we must accept that the only thing which distinguishes the account of these organisms from an account of ourselves is that in the former there occur no private events. As

regards the meaning of the mental terms as used by these organisms we would have to say that, however they function, they don't function so as to refer to anything private, even if they do so refer when we use them.

Now, in Stage II we simply say that we can (and must) consider ourselves to be like the organisms in the example. However 'private sensation' functions, it does not refer to anything private. Nevertheless, it may have a use:

> 'The sense datum is private' is a rule of grammar, it forbids the use of such expressions as 'they saw the same sense datum'; it may (or may not) allow such sentences as 'he guessed that the other had a sense datum of this ... kind'. It may only allow expressions of the form: 'The other looked round, had a sense datum and said ...'. You see that in such a case this word has no use at all. But if you like to use it, do! ((Wittgenstein (3), p. 273.))

So, in a deep sense we are automata:

> 'But surely I know that I am not a mere automaton!' - What would it be like if I were? - 'How is it that I can't imagine myself not experiencing seeing, hearing, etc.?' - We constantly confuse and change about the commonsense use and the metaphysical use. ((Wittgenstein (3), p. 275.))

Of course, we have a use for 'automaton' such that, for instance, 'Barry is not just an automaton' can be correctly used. But in a deeper metaphysical sense Stage II projects the view that we are just automata, in the sense that there just are nothing over and above physical events occurring in the body.[30]

If then the PLA is right, P-primacy is impossible; and if P-primacy is impossible private events could not be distinguishable even by the person in whom they occur, so the notion that they occur would be incoherent. And as we saw, the occurrence of (mental) events in principle imperceptible to others entails at least the Residual Privacy noted in 2.32: - entails at least the possibility of the private distinction between isomorphic private visual imagings (and isomorphic private kinaesthetic sensations etc.). So if P-primacy is impossible there

cannot be events which are in principle imperceptible occurring in a person. And so if the PLA is right we would have to adopt some other notion of the meaning of 'sensation', 'pain', etc., than that they refer to private entities. But the fact that we use these expressions, indeed the fact that there is a use for any expression, is as we see from the organism example above, perfectly compatible with there being no private events. So the fact that we undeniably have a use for these terms cannot seriously obstruct the view that there are only physical events which are in principle observable occurring in us.

If we consider again the organism example, and consider their possible use of statements like 'there are private events occurring in me' - we may find a meaning, a coherent function for such statements, though this is perhaps unlikely. But at any rate, in as much as we do understand such statements we would have to regard them in one sense at least to be false.

We must remember that absolutely any verbal output, any behaviour on the part of these organisms is compatible with only public physical and no private events occurring in them. So any behaviour we observe in any other one of us is compatible with there being no private events occurring in that person (one of us); since there can be no observable difference by which we could distinguish one of us from one of these organisms. But of course if we accepted Stage II in which no private events occur, we would have to rule out any verbal output as having the meaning or implying that there occur private events, and being true. But the important thing to remember is that there is no need to interpret the occurrence of mental ascriptions as implying the occurrence of private mental events, since nothing observable would change if we went from a state in which (a) we did not know that these organisms were purely physical and had no P's, and of course happily ascribed to them anything

we would be prepared to ascribe to ourselves; to the state (b) in which we (somehow!) learn that these organisms have no private events occurring in them, all there is to them being their complex physical states and behaviour.

So, ontological monism is compatible with the sense of all our mental ascriptions except for those asserting the occurrence in us of events which - being additional to bodily and behavioural events - would <u>have</u> to be private to one person and imperceptible to others.

This point needs emphasis, because it means that if we decided that there were no private events occurring in us, although we could continue to use practically all the mental terms we do use, we would rest assured that there was no ontological problem of mind - that ontological dualism was not true. The only problem that would remain concerns how we assign truth conditions to statements containing mental terms, given that they are never used to refer to or describe anything non-material (for there would <u>be</u> nothing non-material). We might tackle this by adopting a <u>linguistic</u> mind-brain identity theory, proposing that some of these terms refer or are used to refer to certain physical events occurring in the body; or we might adopt a more Wittgensteinian account in which expressions like 'my pain' aren't genuine referring expressions: their function in language moves is not one in which they <u>refer</u>. But whatever we decided regarding this linguistic problem, if there were no private events, we could rest assured that there was no deep metaphysical/ ontological dualism of mind and body. Absolutely all the <u>events</u> occurring in us, all our behaviour, including all our langauge moves, would be the sort of events capable of physicalist explanation. For the sort of events which seem to escape physicalist explanation would just not be occurring in us. That's to say, privacy is the only ontologically problematic criterion distinguishing the mental. Other criteria

distinguishing the mental are compatible with ontological monism and the possibility of a physicalist description and explanation of all the events occurring in us and in which we partake. Other mental concepts present no more of an ontological problem than do concepts of arithmetic, chess and so on: no more than the fact that there may be a sense to describing an event or process occurring in a machine programmed to play chess by 'It moved its queen back', 'It's planning a queen-side pawn roller', obstructs a physicalist description and explanation of all the events occurring in the machine, or suggests that there is an ontological dualism involved.

Even if we did perversely decide that the organisms in the example could not properly be described as having desires, intentions, feelings, and so on, it could only be because there were no private mental events occurring in them, not because they were different in any other respect (indeed, different in any observable respect). We would otherwise have been perfectly happy to describe them in this way. That's to say, if the ascription of desires, thoughts, feelings and so on is taken as entailing an ontological problem, this could only be so by taking these mental concepts as slaves of privacy. Whether or not there occur private mental events is the crucial issue for the ontological mind-body problem. Everything else mental is either compatible with ontological monism, or, if deemed incompatible, only in that it would be taken to require the occurrence of private mental events.

Thus, at least for the weaker form of physicalism which maintains that there is in principle a physicalist explanation for every event occurring in a human being or in which a human being partakes, the question of the occurrence of private mental events presents the only logical obstacle to physicalism. For only ontological dualism can logically obstruct physicalism of this form, and ontological dualism requires

that there occur in people events which are additional to any events which are in principle perceptible; which in turn requires that these additional events are private to the person in whom they occur. And if physicalism in this form is sufficient as well as necessary for the stronger form which maintains that the human mind presents no logical obstacle to its scientific accommodation, then the question of the occurrence of private mental events presents the only logical obstacle to physicalism of the stronger form.[31]

We may summarise these considerations as follows:

First, we assume (a) that all physical events are either in principle perceptible directly to our senses, or in principle perceptible indirectly in that we can conceive how they could causally engage with events we can perceive directly. (We take an event satisfying this disjunction to be in principle perceptible; and an event satisfying neither disjunct as in principle imperceptible.) And (b) the only events which are in principle perceptible are physical events. (It may be objected that we can perceive or observe events which are not physical: for instance, that we can perceive or observe someone in pain, but that what we perceive then is not necessarily a physical event. Certainly, in the sense that we can perceive or observe that someone is in pain we may not be perceiving or observing that a physical event occurs. But in the extensional sense in which we would be directly perceiving an event when we perceive someone in pain - say, writhing about - the event we perceive is physical: it has a physical description (at least in principle). Whatever particular dated event we visually (or otherwise) perceive we must construe as having (in principle) a physical description. In the sense that we may perceive something-which-isn't-physical what we perceive cannot be an ontological rival to the (physical) event we thereby perceive. For either what we perceive cannot be construed as a

particular event; or, if it can, it must in principle have a physical description. Thus it may be argued that "a wedding" isn't physical; but in the sense that we are perceiving an event when we perceive a wedding, we are perceiving an event which has a physical description.)

Secondly, any event occurring in a person which is in principle imperceptible to others must be such that that person nevertheless has access to it of a sort - otherwise the supposition that such an event occurs immediately loses coherence, for no one could then have evidence for its occurrence. Now, we must construe this access as logically privileged access. For (a) the person must have criteria for recognising the occurrence of such an event, criteria of a sort which only that person can have, since other people can only base their evidence indirectly on events which they can perceive; and (b) other people could only check the accuracy of their evidence against the evidence the relevant person has: since other people could not, for example, check that a certain event which is necessarily imperceptible to them occurs every or any time their evidence suggests it does - except by taking into account the verdict of the person who is thereby privileged. Thus we must construe the relevant person's access as logically privileged, and requiring the possibility of P-primacy. Moreover, in the case where the events in question are visual imagings, there is something in principle incommunicable - Residually Private - about such events ; again, in requiring the possibility of a necessarily private distinction by the relevant person, requiring the possibility of P-primacy. In other words, any necessarily imperceptible event occurring in a person must be private (in a sense obnoxious to Wittgenstein).

So any event occurring in a person which was not physical and occurred in addition to all the physical events occurring in that person, would be in principle imperceptible, and hence (in principle) private.

And of course, no private event occurring in a person could be physical.

Now:-

(i) The occurrence of absolutely all our behaviour, including all our linguistic outputs and so on, would be compatible with there only being physical events occurring in us (and no private mental events); and is therefore compatible with the possibility of there being a physicalist description and explanation of all the events occurring in us.

(ii) On the supposition that it is coherent to allow that private events could occur, we could conceptually distinguish between ourselves and organisms physically indistinguishable from us, but such that the former organisms present no ontological mind-body problem: only physical events occur in them. But the only and crucial feature which would at least conceptually distinguish us from these organisms is the occurrence in us of events which are in principle imperceptible (publicly) and are therefore private. In other words, the only way in which we could be taken to require an ontological dualism of events, while these physically indistinguishable organisms don't, is if there occur in us (but not in these organisms) private events. So the occurrence of private (mental) events is necessary and sufficient for there to be an ontological dualism of mental and physical events.

(iii) If the PLA of Wittgenstein's is successful, we cannot coherently suppose that private mental events can occur. So there could not then be supposed coherently to be an ontological dualism of mental and physical events. So there could not then be a logical obstacle to the weaker form of physicalism. (Nor could there be coherent "other minds" scepticism; and of course, the distinction made in (ii) could not then be coherent.)

I will refer to (i) as the Materialist Compatibility Point. (It

is of course inherent in any dualist view, such as the parallelism of Geulincx, which nevertheless allows that there is in principle a physicalist explanation of all bodily events.) Naturally it is based on an assumption that a purely physical structure could in principle do all we do. But I do not know of any sound arguments ruling out this possibility.[32] The Materialist Compatibility Point is sometimes invoked without recognition that point (ii) is a consequence of it. (We shall see this later in the case of Davidson.)

Point (iii) is frequently unrecognised. In particular, there is a tendency in some philosophers to fall between two stools with regard to the PLA, and suggest that all the PLA need show, if it is right, is that we can communicate everything about, for example, what sensations we have, or more generally, what mental events occur in us.

But it is often then left in the air how we are to construe mental events - or indeed, whether there are mental <u>events</u> which we can be introspectively aware of. Are mental events still to be construed as necessarily imperceptible? Is the having of a sensation still to be construed as a necessarily imperceptible event?

If the answer is <u>yes</u>, then we simply must suppose that such an event is private, and occurs in a "medium" to which the relevant person has logically privileged access. So we cannot accept the full consequences of the PLA and give this answer.

And if the answer is <u>no</u>, then it must either be supposed that we cannot be introspectively aware of any mental <u>events</u> - in which case it is very difficult to justify <u>ontological</u> dualism of mind and body; or, if mental <u>events</u> are allowed, then since they are in principle perceptible, they must in principle have physical descriptions. And again ontological dualism cannot be sustained. If we do allow mental events of which we can be introspectively aware (and the PLA does not rule this out -

though Wittgenstein can be interpreted as taking the line that there are no such events - it would only show that such events could not be (in principle) private), then there may still be a problem regarding what are the right physical descriptions for such events. But it is often overlooked that if the PLA is right and we allow such mental events, then there must in principle be a physical description for the latter.

The commitment to one of these answers, and an acknowledgement of the consequences for ontological dualism, are frequently avoided by those who accept the PLA. But as a consequence of the PLA only the negative answer is coherent.[33]

The sort of prevarication that can go on here regarding the import of the PLA is not really surprising in the case of someone who:

(a) Supposes that the PLA is right and that there are no private mental events.

(b) Supposes that he can be introspectively aware of mental events occurring in him - only they're not private in principle.

(c) And supposes that his mental events do not and cannot have physical descriptions: cannot be physical events. In particular (he supposes), they are not the events other people perceive when they obseve his overt behaviour; nor are they the events other people could perceive if they observed his brain working.

For such a person can only follow (a), (b) and (c) (or follow (a), (b) and abstain regarding (c)) if he does neglect to consider carefully:

(d) That if his mental events are not in principle private, they must in principle be perceptible to others, and therefore must have physical descriptions. (For if in principle they were none of the events other people could perceive occurring in him, then other people's evidence for their occurrence could only be indirectly based on events they could perceive; in which case other people would have no way of

checking whether they occur or not except through the privileged evidence available to him. His evidence is then not only of a different sort, unique to him (not necessarily by itself in conflict with the PLA, as we shall see - it is only in conflict if his evidence is logically superior), it is also logically privileged.)[34]

Sometimes, because Wittgenstein is able and willing to say that there is a sense in which there is a distinction between, say, pain and pain-behaviour, point (iii) is unnoticed as a consequence of his views. But Wittgenstein is simply admitting that 'pain' and 'pain-behaviour' have a different use in language moves. Nevertheless, there can be no events additional to those which are in principle perceptible.

Glover says:

> Neither in the view of Wittgenstein nor in that of Strawson can mental states be analysed without remainder into behaviour. These theories are of exceptional interest, but they too leave us with problems. No completely clear account of the precise logical links between mental states and their outward manifestations on this kind of view has yet appeared. And, in so far as mental states are not completely analysable in terms of behaviour, the old question, if it is a genuine one, of what exactly mental states are seems still to arise. ((Glover, p. 8.))

But in Wittgenstein's view, although there may be problems stateable in the form: 'What exactly does "sensation" mean?',etc., we know that whatever 'sensation' means the problem is one of meaning which, it is suggested, cannot spread into a problem for the physicalist. Assuming there was a full physicalist theory of the physical events occurring in peoples' bodies, there wouldn't in Stage II still be a special problem for the physicist regarding the meaning of mental concepts.[35]

In contrast Strawson does not eliminate the classical mind-body problems: on his account there does remain the possibility of there being bodily states and behaviour and something else. This is particularly evident in his discussion of the possibility of a disembodied person ((2), pp. 115 - 116): - for Strawson it is conceivable that a

person can survive disembodied after death, though it is not intelligible for Strawson that a person could conceive of himself as a person, in a disembodied state, unless he was once embodied. But it is quite sufficient for the survival of mind-body problems for the physicalist that there is something to an embodied person which could survive disembodied.

And in contrast to Ryle, Wittgenstein's ammunition against dualism is such that there is no need for him to attempt a detailed linguistic analysis of mental terms for the metaphysical success of his Stage II behaviourism. And Ryle's view suffers because of the absence of any effective argument of his to render impossible the occurrence of private events in a person. Provided their occurrence remains a possibility and we can insist that 'my having a private sensation of red' can be used to refer to a private event, we can insist that private mental events and physical events belong to the category of events (occurring in time), and thus reinstate ghostly events additional to physical events.

Certainly many mental concepts can and may have perfectly good sense without bringing into that sense anything which is in principle unobservable. (This can be construed as a consequence of the Materialist Compatibility Point.) And that sense may place these concepts in a different category from physicalist concepts, just as chess concepts are in a different category from electronic concepts. And Ryle may succeed thus in showing that these mental concepts need not trouble the physicalist. But it is really private mental events that cause the trouble, and whereas Wittgenstein goes beyond the Material Compatibility Point and tries to prove that there could not be a coherent distinction between us and the organisms in the example, Ryle does not confront this sufficiently to achieve his exorcism.

Hopkins expounds quite clearly a Stage II position: he makes what

is essentially the Materialist Compatibility Point, by means of an example of a possible world just physically exactly like ours, and concludes, explicitly accepting the PLA as successful, that 'for things to be truly describable as we describe them, they need only be (physically) as they are.' (p. 142). Thus Hopkins finds the way clear for physicalism essentially as a result of the elimination of private entities by the PLA, (which as I say he accepts). And he also suggests what is again implicit in Wittgenstein, and also related to the Davidsonian line, that it doesn't matter for physicalism that 'expressions from the vocabulary of physical science differ in meaning, or lack useable expressive power or focus, as compared with those in humane use.' (p. 143).

For observable things to be truly describable as we describe them, they need only be physically as they are. But are things only physically as they are? Hopkins's point carries only if there are no private (mental) events - if the PLA is successful.

2.5 The Possibility of P-Primacy

The possibility of a private language and the coherence of the notion of P's devolve upon the possibility of a language user's being bound by P-primacy. Under this condition only the one user can be acquainted with the criteria by which the expressions in the language mean what they mean.

The sense of "private language" involved is of course such that the language is in principle rather than contingently private: the unintelligibility of the language (to others) is logically guaranteed in virtue of what its sentences say, in particular, what its referring expressions refer to. Thus Wittgenstein says:

> The individual words of this language are to refer to what can only be known to the person speaking; to his immediate private sensations. So another person cannot understand the language. ((I, 243.))

It is important to bear in mind that putting the issue in the form of the possibility of a private language is a convenient way of handling the intimately connected possibilities of private language and private recognition.

This helps to bring out the point that one of the things Wittgenstein is anxious to show is that the mechanism by which we use public natural language is not one which contains the implicit requirement of the possibility of P-primacy. For in Stage I we use public natural language all right, but contained in the account of how we do this is the requirement that P-primacy is possible - that there is a private component underlying the public agreement in natural language. This is sometimes overlooked. Hervey, for instance, asks whether in a situation where Crusoe is supposed to have private sensations, he and Man Friday could understand each other; and concludes that Crusoe could teach Man Friday the meanings of the words of his sensation language:

> But if, as was suggested before, he had impulses to perform certain actions when he experienced his sensations, and if these actions that he wanted to carry out, but did not, corresponded to Man Friday's actions, then he could work out a correct correspondence between his words and Man Friday's. When he had done this he would be able to teach Man Friday the meaning of the words in his language, at least of those which referred to the sort of sensations which Man Friday had also. ((Hervey, p. 91.))

But underlying the agreement in use of words here, is the implied possibility that there is first a private recognition of private sensations. This is essentially the situation in Stage I: there is communication; but it is not the meaning of terms bound by P-primacy that is communicated. And the Residual Privacy attending this sort of set-up means that something will remain incommunicable.

It is also important to note that the crucial point that would separate a public language from a private language is not that in the former other people are involved. The desert island factor is irrelevant in that (a) neither a solitary isolated person nor a person in a

community could, according to Wittgenstein, recognise or label private sensations; and (b) a desert island linguist could for various reasons - as a diversion, an aid to memory, a record for posterity - develop a language. But according to Wittgenstein that language would in principle be intelligible, because the criteria of meaning operating in it would be in principle accessible to anyone.

We can easily specify the contrast wittgenstein wants between permissible conditions governing the use of language and the impermissible P-primacy, without requiring more than one person to use the language. Essentially the contrast depends on whether or not there are genuine criteria governing the use of the language: -

Suppose that I decide to label a colour in the following way. By a certain patch of colour on a piece of paper I put the sign 'S'. Suppose now that I want to check what 'S' stands for. But when I look at the piece of paper with the colour-patch next to 'S' I think that originally when I wrote 'S' down the patch was coloured differently. If in this case I nevertheless let the reference paper be the final arbiter, and call by 'S' whatever colour the patch is when I check, then I am constrained by a genuine criterion.[36] And of course my use of 'S' then is in principle intelligible to others: - someone else could come along and also inspect the reference paper, and in principle understand my use of 'S' - the reference paper is a public object (at least in principle).

But suppose I wanted to use 'S' as a label for my private sensation of the colour of the reference patch, for my P(v something the colour of the patch), (which I will abbreviate to 'P(v patch)'). Now in this case I could not use the reference patch to provide a criterion for using 'S', for if when I go to check the patch, my P(v patch) seems to have changed from the original, S, I couldn't carry on and label the new P(v patch) by 'S' also - for then I wouldn't be labelling the original private

sensation: the original private sensation would not then have fixed the meaning of 'S'. At best 'S' would then be a label for whatever my P(v patch) happens to be when I check. And suppose it's thought that I could distinguish between merely thinking or having the memory that the original P(v patch) was different, and the original P(v patch) actually having been different: - that in other words, my memory is right. Well, how could I make this distinction? It's no good my appealing to anything else - because whatever anything else tells me it's possible that my memory of the original P(v patch) was right all along, only that P(v patch) has changed and nothing else has changed. If I go and look at patch 2, remembering that the original P(v patch 1) was the same as P(v patch 2), and find that P(v patch 2) doesn't seem to have changed, so then I decide my memory of P(v patch 1) is wrong, I am then failing to incorporate the possibility that P(v patch 1) changed but P(v patch 2) didn't. And if I fail to incorporate this possibility I am failing to label my individual private sensations.

We can see how this is going: if I use anything as a check or criterion for whether or not my P(v patch) has changed, other than my memory that it was S and now isn't, then the original private sensation is going to get lost. My original private sensation can only fix the meaning of 'S' if nothing but my memory of it, what I think it was, serves as a criterion. For everything but my memory of it, of what P(v patch) was, is logically independent of what my P(v patch) was - nothing else seems to depend on what my P(v patch) was. So if anything else serves as an overriding check, overriding my memory or what I think it was, then this other overriding rule will fix the meaning of 'S', and not the original private sensation: - my P(v patch) could have changed even though the overriding rule makes me continue to use 'S'.[37]

Now in this case of course my use of 'S' would be unintelligible

to someone else who comes along: for the public criterion he can observe as possibly fixing the meaning of 'S': - the reference patch - is not governing my use of 'S'.

It is important to note that we don't want to say that when the reference patch provides the criterion for using 'S' we are then bound by B-primacy, which might suggest that Wittgenstein's arguments may merely show that we must always be bound by B-primacy. For B-primacy is merely the Stage I device for construing communication between men: the device by which we construe "common ideas" among individuals each having "private ideas". Both P- and B-primacy, as I outlined them, require the possibility of private recognition - they are embedded in Stage I.

What we can observe though is that whether the criterion for using an expression is given by something like the reference patch or by other people correcting our mistakes doesn't matter in characterising the use of the expression as public (at least in principle). What matters is the sort of criteria we use: if the meaning of an expression is governed by something privately accessible to only one person, then the language is in principle private. And Wittgenstein argues to undermine the possibility of there being genuine constraints on private use.

2.51

> Let us imagine a table (something like a dictionary) that exists only in our imagination. A dictionary can be used to justify the translation of a word X by a word Y. But are we to call it a justification if such a table is to be looked up only in the imagination? - "Well, yes; then it is a subjective justification." - But justification consists in appealing to something independent. - "But surely I can appeal from one memory to another. For example, I don't know if I have remembered the time of departure of a train right and to check it I call to mind how a page of the time-table looked. Isn't it the same here?" - No; for this process has got to produce a memory which is actually correct. If the mental image of the time-table could not itself be tested for correctness, how could it confirm the correctness of the first memory? (As if someone were to buy several copies of the morning paper to assure himself that what it said was true.) ((I, 265.))

> I can look at the clock to see what time it is: but can I also look at the dial of a clock in order to guess what time it is; or for the same purpose move the hand of a clock till its position strikes me as right. So the look of a clock may serve to determine the time in more than one way. (Looking at the clock in the imagination.) ((I, 266.))
>
> But in the present case I have no criterion of correctness. One would like to say: whatever is going to seem right to me is right. And that only means that here we can't talk about 'right'. ((I. 258.))

Wittgenstein's argument is essentially that there can be no genuine criterion for saying whether two P's are the same or different, because under conditions where we would be trying to identify or re-identify a particular P, thinking that one was right would entail being right. And if the notion of 'right' is to have sensible application, this must not be the case: it must not be the case that one could never be wrong.

We can produce more examples. There can be criteria for telling if we have hit or missed a target we are aiming for; but what would be the criterion if the target was given as whatever we point our rifles at? We could surely not then say that we were aiming, or that what we hit were targets. The point is this: in this situation we could not be constrained by the proposed criterion or instruction: 'your target is whatever you aim for'. What would we do when instructed to shoot according to this instruction? Well, suppose we did point at something and shoot: we didn't then shoot that because of the instruction, because of the criterion given for what was to be our target. For we could have shot at anything and conformed with the criterion. In other words, the criterion does not separate targets from non-targets. Whatever we did shoot at, supposing we shot at all after this instruction, would have to have been determined by something other than that instruction - for instance, we might say: 'We might as well shoot rabbits' or 'Let's spin the rifles and shoot where they end up pointing'.

Again, adapting Wittgenstein's example, suppose a clock-maker was

instructed to build a clock so that whatever the hands say we will accept as the right time. The clock-maker, if he built a movement for this clock, could not have built that movement because of our criterion - for any movement he'd constructed would have satisfied our demand.

Hopkins has a nice example: someone could not 'claim his movements were constrained by the condition that he had always to go where his shadow might follow' (p. 136).

In all these examples it is impossible to violate the criterion or rule given.

And now imagine that I am trying to use P-primacy to label a particular P by 'S'. What if I had a doubt as to whether that P was labelled by 'S'? But what's the point of having a doubt when I know beforehand that whatever I decide will fix the referent of 'S'? "Whatever I decide" will not furnish me with a criterion for using 'S' in a way that it would mean anything - that it had restricted application. So I could not be labelling a particular P. But suppose that I insisted that my constraint for labelling a particular P by 'S' was not "whatever I decide", but my memory. But then, so the argument goes, it would not be the case that that particular P fixed the meaning of 'S', because, as Wittgenstein says of the P: 'assume that it constantly changes, but that you do not notice the change because your memory constantly deceives you' (I, part II, p. 207).

2.511 Objections to the PLA tend to fall into two broad categories. The first, the most frequent sort of objection, is that contrary to Wittgenstein's contention there is a sense in which rules could be followed in a private language - there is a possibility in such a language of criterial constraint. The second sort takes Wittgenstein's argument as based on an extreme scepticism, essentially a scepticism about memory, which would have the consequence that we would be unable

to follow any rules at all.

Objections of the first sort often misinterpret what in this context constitutes a private language, or else find a notion of checking which when implemented results in the language for which these checks supposedly furnish criteria not being a private language.

Hervey for instance tends to misinterpret what constitutes the privacy of a private language, and assumes that a language in principle intelligible to only one person must be ipso facto devoid of rules:

> Does this mean that it must not in principle be understandable, that is to say, that it is logically impossible for someone else to understand it, or does it simply mean that he could not understand it in practice? If the former is meant, then this would appear to amount to the demand that the 'language' shall be rule-less or have no regularities whatsoever, either of usages or constructions... The only sensible interpretation of the demand, then, is to interpret it as meaning that the language could in principle be understood by someone else, but that it could not in practice be understood unless its inventor were in a position to explain it or provide clues for understanding it and he chose to do so. ((Hervey, p. 85.))

But the issue concerns a language in which there would be regularities, but these regularities are intelligible only to one person because they link up with a private substratum. Thus Hervey's subsequent attack fails to engage with the issue Wittgenstein was concerned with.

An example of the failure to notice that the suggested implementations of checks preclude the operation of P-primacy occurs in Ayer:

> But if we allow that our Robinson Crusoe could invent words to describe the flora and fauna of his island, why not allow that he could also invent words to describe his sensations? In neither case will he be able to justify his use of words by drawing on the evidence provided by a fellow creature: but while this is a useful check, it is not indispensable. ((Ayer, p. 56.))

Ayer then argues that though Crusoe's mistakes may not make any practical difference to him, this does not mean they are not mistakes.

Two points need again to be emphasised here. The first is again that other people are not a crucial factor - the possibility of having genuine criteria for the use of words does not require fellow language

users. The Crusoe setting is a red herring. Secondly, Wittgenstein points out that the sort of checking that Ayer feels is possible for Crusoe or any supposed private linguist, would mean that P-primacy would cease to operate:

> I discover that whenever I have a particular sensation a manometer shows that my blood-pressure rises. So I shall be able to say that my blood-pressure is rising without using any apparatus. This is a useful result. And now it seems quite indifferent whether I have recognised the sensation right or not. Let us suppose I regularly identify it wrong, it does not matter in the least. And that alone shows that the hypothesis that I make a mistake is mere show. ((I, 270.))

We might try and retort: imagine that I forget which private sensation accompanies the rise in my blood-pressure, but luckily I remember the association of 'S' and 'T' with particular private sensations. I look in my diary and see that I've written 'S occurs when my blood-pressure rises' - this now tells me something, for I can remember which private sensation 'S' stands for. Moreover, when subsequently I have the private sensation S, and my blood-pressure rises, I thereby check my annotation. But if this is a genuine check it precludes the operation of P-primacy regarding the use of 'S': the same considerations apply that attended the reference patch example. For if my diary annotation provides a genuine check, a genuine constraint on use, then it is not a particular private sensation that fixes the meaning of 'S', but the rising of my blood-pressure: - for if on one occasion I think a different private sensation accompanies the rise in my blood-pressure, not S, then if I nevertheless accede to the check and label this private sensation 'S', then 'S' is not labelling any particular private sensation, for any private sensation that had occurred then I would have labelled 'S'. And if I override the "check" and decide it's a new private sensation, not S, which on this occasion accompanies the rise in my blood-pressure, the "check" is no check at all: my diary annotation did not constrain my use of 'S'.[39]

Mundle misses this point:

> Suppose our diarist is liable to suffer from spells of tummy trouble which his doctor cannot diagnose. The onset of such attacks has been preceded by distinctive shooting sensations in his midriff... Our diarist starts writing 'E' in his diary ('E' being short for 'those damned twinges') in order to verify whether E's are <u>regularly</u> followed by the other symptoms, because this information might help the doctor's diagnosis... It is surely unwarranted dogmatism to say that in such cases 'E' has no meaning for the diarist. ((Mundle, p. 110.))

And also Castaneda, who argues that a private linguist speaking the private language Privatish 'can resort to practically all the "things" to which speakers of public, or of ordinary, languages have recourse - he can do this even in the case in which Privatish is a <u>purely</u> private language.' (Castaneda p. 147).

As I mentioned above, the second broad category of objections to the PLA interpret it as involving scepticism about memory. If we can never in general be sure whether our memory is reliable, then we can never be sure if we remember the use of any sign right, even in natural language - and yet we don't conclude that natural language is not language. Thus Ayer argues that verification must have a stop somewhere (p. 57) , and that if scepticism is effective in the PLA, it would have as much effect against ordinary language. But, as Kenny points out (p. 217), the PLA does not rely on <u>scepticism</u> about memory. Wittgenstein is not arguing that we can never in general be sure whether our memory is reliable. Rather, he is arguing that while the concepts of good and bad memory have application to the use of signs in ordinary language, there would be no room at all for these concepts under P-primacy. One might doubt whether one had remembered right in ordinary language; but here there is a sense to "remembering right". But there is simply no sense to "remembering right" under P-primacy. In the case of the reference patch example, when I'm using it as a genuine check I have a criterion for the correct memory: if I remember the patch to be a

certain colour, I can check my memory: it's right if it is the colour I remember, wrong if it isn't. But as we saw, under P-primacy I have to accept every memory as "right", or every decision on the basis of memories as "right" - I can't misremember under P-primacy, for I cannot appeal to something else to check my feeling that my private sensation has changed.

A more subtle form of the memory objection is made by Pears, who argues that:

> ... the speaker's memory can be checked when he makes statements about material objects, and then, having passed this test successfully, it can be used without any possibility of a check, but nevertheless with reasonable confidence of regular results, when he makes statements about his own sensations. ((p. 161.))

But what if I had a very bad memory normally when using ordinary language, and I constantly made mistakes? Would that be an indication that I shouldn't trust my memory when, as under P-primacy, I can't check it? An undesirable result of this is that whether my private "language" was a language might depend on how good my memory contingently happened to be. But the main difficulty with this objection is exposed by Wittgenstein:

> "Imagine a person whose memory could not retain what the word 'pain' meant - so that he constantly called different things by that name - but nevertheless used the word in a way fitting in with the usual symptoms and presuppositions of pain" - in short he uses it as we all do. Here I would like to say: a wheel that can be turned though nothing else moves with it, is not part of the mechanism. ((I, 271.))

In other words the "efficacy" of P-primacy memory is independent of the efficacy of my memory in ordinary language. However good my memory is when I have criteria for remembering right, it simply doesn't engage with the "efficacy" of my memory under P-primacy. For it seems possible, as Wittgenstein envisages, that I might constantly misremember which private sensation 'S' labels, and yet fortuitously I use ordinary language right. So I would have to make the further assumption that

my memory for private sensations was "good" as well as (not because of) being good when remembering ordinary things. And now what sense is there to this further assumption? Because I cannot have criteria for remembering my private sensations right, there is no sense to comparing how good my memory for private sensations is with how good my memory for ordinary things is. We must accede to the fact that we can have no notion of correctness or reliability for memory of private sensations.

2.512 ... to think one is obeying a rule is not to obey a rule. Hence it is not possible to obey a rule 'privately': otherwise thinking one was obeying a rule would be the same thing as obeying it. ((I, 202.))

Thomson challenges this contention. She puts this contention like this:

> A man's use of a sign is not governed by a rule unless it is not merely possible that he should violate the rule but more, that he should violate it unwittingly. That is, it must be possible that he should think he is following the rule, it need not follow that he is really following it. ((p. 189.))

She then tries to produce counterexamples to this contention:

> ... consider the rule 'Always decide to do what you think at the time it would be most fun to do'. Note, not 'Always do what etc., etc.', for I might think I was following this rule, and not really be following it - e.g. all unknown to me I am paralysed, and so I am not doing what I think I am doing. ... Is there anything in the concept 'rule' which rules out that this should be a rule? ((p. 190.))

> Or again, Mummy writes, 'Whenever you feel the least bit gloomy, think of your Mummy. (It will cheer you up.)' Could I think I was following this rule and not be? ((p. 191.))

Certainly, if one is thinking of Mummy this entails one is thinking of Mummy. But thinking that one is thinking of Mummy (thinking one was obeying the rule) does not entail that one is thinking of Mummy. One can be mistaken about what it is to think of someone, or one might be mistaken about who Mummy is. That's to say, the criteria for using 'I am thinking of Mummy' correctly are such that thinking one was observing these criteria in asserting that sentence does not entail that one is observing them. The same consideration applies to Thomson's other rule.

There are many ways in which I might think that I can truly say of myself 'I have decided to do what I think it would be most fun to do', when in fact I am wrong in thinking that.

Nevertheless, a case may perhaps be made from Thomson's view that one may question as evident that 'A man's use of a sign is not governed by a rule unless it is not merely possible that he should violate the rule but more, that he should violate it unwittingly'. (My underlining.)

2.52 We can put the notion that a criterion is genuine in two related ways. The criterion must be able to function as a constraint. And, what amounts to the same thing, the criterion must provide the means to distinguish what does and what doesn't conform to the criterion: what breaks and what doesn't break the constraint.

Thus 'Go where your shadow follows', 'Aim at where you point your rifle' do not provide genuine constraints.

In the case of a genuine criterion for how 'T' is to refer, the criterion must constrain the use of 'T': it must distinguish correct from incorrect statements of the form '"T" refers to X', or 'T occurs'. And if the only way a specification for the use of 'T' restricts the referent of 'T' is by specifying the eligible referent as that which is referred to by 'T', then we have no genuine restriction at all. With this and the other bogus rules there is no logical distinction between following the rule and following the rule right. Going anywhere, aiming at anything, referring to anything would conform with these bogus rules.

And now it would appear that the only property of my private sensation S that is constant enough to "constrain" my referring to it with 'S', is that I refer to it with 'S'. And this is no constraint.

Now let's consider the following argument against the PLA.

It is supposed, in this argument, that there is a principle which we may regard as analytic and evident which Wittgenstein uses in connection

with the PLA - but that the success of the PLA depends not on that principle, but on one which is by no means self-evident.

If we examine what we have to say about the significant use of a sign, we can extract the following analytic principle from the requirements for a criterion of the relevant sort to be genuine:

(i) If the term 'S' is going to be significant in a language, there must be an effective criterion by which a distinction can be made to separate correct from incorrect uses of 'S'. So if 'S' is an appropriate referring term, the criterion will separate true from false statements of the form 'S occurs'.

Now consider:

(ii) Not any use of 'S' can be right. There must be a sense, given by the criterion, in which the use of 'S' is wrong or mistaken.

And:

(iii) It must be possible for me to use 'S' incorrectly.

(iv) It must be possible for me to use 'S' incorrectly while under the impression that I used it correctly.

(v) It cannot be the case that if I think I'm using 'S' right, ipso facto I am using 'S' right.

Point (i) seems evident enough. And (ii) is a respectable consequence of (i). But (iii), from which (iv) and (v) may be supposed to follow, is open to two readings, only one of which entails (iv) and (v); and on that reading it is not a consequence of (i) or (ii). For if (iii) is taken as:

(vi) It must be possible for me to distinguish correct from incorrect uses of 'S' - there must be sense to the distinction.

then it does not entail (iv) and (v). It only has these as consequences if it is read as:

(vii) It must be possible for me to make the distinction between

correct and incorrect uses of 'S', and it must be possible to make that distinction incorrectly.

But only the first conjunct of (vii) is implied by (i).

Of course (it is supposed in this argument), (v) will normally apply in natural language simply because the criteria deciding correct from incorrect uses are not given by me - I have not, no one person has, final authority over the use of expressions in natural language. But where only I have access to the distinction between a correct and an incorrect use of a sign, (v) will be breached, but (i) may be fulfilled.

It may be argued then that in fact Wittgenstein does not show that condition (i) is not fulfillable under P-primacy; though it is admitted that (v) is not fulfilled. But then (v) does not seem to be an analytic restriction on there being sense to a sign in language - though it would be an analytic restriction on any language over which I would not have final authority - such as public natural language.

Thus it would be argued that (i) and (ii) are fulfillable under P-primacy simply because I could distinguish correct from incorrect uses of 'S': if 'S' is a term referring to a private sensation, I could separate true from false statements of the form 'S occurs now', 'S occurred then' - for if I think or remember that S occurred just then, I cannot accept as true that just anything occurred just now, that T and not S occurred just now. Even though I can't remember right. But, it would be argued, I don't have to remember right for my memory to be a genuine constraint on my use of 'S' - I can't after all choose to believe or remember just anything. So even if there's no sense to my believing, remembering correctly under P-primacy, I am constrained by what I do believe and remember. And then I can ipso facto conceive of wrong uses of 'S', even though I can only use 'S' wrongly knowing that I use it wrongly.

2.53 But then the sort of entity I am referring to with 'S' is such that it cannot be logically independent of my memory or belief "that S occurred". It might be supposed that this is exactly what we must expect of private entities - that they are somehow logically tied to what we believe, remember and so on. But then it becomes terribly difficult to make sense of the claim that they are genuine entities.

Thus consider the case where it is supposed that I privately recognise a certain private mental event as occurring. Clearly, the occurrence of this event cannot now be logically independent of my having a certain thought, or memory, or "act of recognition". But then we cannot make sense of the claim that it is an event which is distinct from the memory, or "act of recognition". For it cannot be distinct from these and logically dependent on them. And if it is not distinct from the relevant memory or belief, then even if we could continue to construe sensibly 'a belief that S occurred' - which we could not - S just could not be construable as an event in any way distinguishable from having the relevant memory or belief. But it is not the having of a belief or memory that is supposed to be being referred to by S, being privately recognised as S: the belief or memory is supposed to be providing a constraint for the recognition of something other than the belief or memory. In order to recognise a genuine thing or event, there must be the thing (or event) and the recognising of it.

Thus:

> Always get rid of the idea of the private object in this way: assume that it constantly changes, but that you do not notice the change because your memory constantly deceives you. ((I, pt. II, p. 207.))

If it is a genuine "object", then we have to allow that it can occur logically independently of the memory. If it cannot occur logically independently of the memory, it is not a genuine object and the memory can't be a memory of it.

Consider the following analogy.

We can allow a sense in which a person may have final authority about something: thus, suppose we are playing "follow my leader", and I am the leader, and the games are to use terms like 'S' to label things. I have final authority on the use of 'S', 'T' and so on: I cannot be said to mislabel when I start the game off.

Similarly, suppose the "follow my leader" involves me (the leader) shooting at something with my rifle - what I hit henceforth becomes the target for others. Again I have final authority. When I shoot I can't be said to <u>miss</u>.

But in these cases there is a sense to saying that I label <u>something</u>, shoot at <u>something</u>. The thing I label, the thing I shoot at are such that they <u>can</u> be specified independently of the fact that they are to be labelled or shot at. I may shoot a tin can: but the tin can isn't the sort of thing for which I <u>must</u> have final authority in electing it as a target. We <u>can</u> describe it independently of my shooting <u>at it</u> - in spite of the fact that I couldn't have been said to shoot it "wrongly". Certainly, in the context of the game I had to succeed, whatever I hit (so there was no "success" about it). But then whatever I did hit isn't the sort of thing I would always <u>have</u> to succeed at hitting whatever I shot at.

The analogy for the case of attempted private recognition is this: it would be rather like supposing that I could be said to shoot the barrel of my rifle (or perhaps, that I could be said to hit the bullet). These are the sorts of things which I would always "have to succeed at hitting", and therefore not the sort of things I could be said to hit.

Trying to recognise something, like trying to shoot something, are cases where we need a certain division between two "things" to make sense of the activities. We need some division between a subject and the

thing he recognises, between the bullet and what is hit. Thus the opponent of the PLA supposes there is a sense to envisaging a diverse private environment which he, "standing back", can recognise, identify and so on. But where is the "he" that is recognising these private items?

This brings us to a rather more global point which the PLA can be interpreted as making. This global point, involving a major difficulty concerning the Stage I view, is perhaps familiar in contexts not always associated with the PLA. But it is intimately connected with the PLA - indeed, concentrating on the minutiae of the private language controversy may obscure this perhaps most salient feature of the PLA.

I will make this more general point; and then consider a dilemma which wittgenstein's line seems to leave us with.

2.6 The Banishment of Private Mental Events and the Problem of Introspection

There is a very important general difficulty with Stage I which arises through the unsound license to extrapolate from contingent to logical privacy. This difficulty involves the notion of a subject of private experience.

The notion of private mental events seems significant provided we remain able to maintain a sense in which, though such an event would be in principle imperceptible to all but one person, we can at least allow that one person has access of a sort to such an event. For if such an event wasn't even accessible to one person, then surely such an event, like a necessarily imperceptible pink elephant (perceptible to no one), would be banished completely.

The point about allowing that something is contingently accessible to just one person, is that we are able to make sense of there being the person and the thing. Supposing that something in principle accessible to just one person can be made sense of by simple extrapolation from the contingent case, is of course the sort of confusion which clouded some of

the arguments cited against the PLA.

Consider the case of the contingent-privacy model I gave for Stage I - the chess model. In this we stipulated a "person", A, and his contingently private environment. While we can construe the contingently private environment as comprising genuine objects, distinct from each other, distinct from A, and so on, we are on safe ground. We are clearly construing A and his environment in the chess model on the lines we construe the normal case of a person and the world, with the difference that in the chess model part of the world was contingently private to A.

We can interpret the PLA as showing that the extrapolation from contingent to necessary privacy precludes us from continuing to construe necessarily private experience by visualising a subject who is aware of his private experiences. A subject can no more be envisaged as aware of objects or events logically tied to him than a man can take a pot-shot at the bullet he fires. So we cannot coherently entertain the notion of a person who is aware of or has access to or evidence of objects or events which are in principle private to him. The notions of "subject", "person", "object", "event" do not survive the extrapolation. We cannot separately give significance to 'A' and 'A's private sensation S'.

Thus we cannot construe a necessarily private environment as a repository of objects of which a subject is aware.[40] But then to whom can private mental events be private? Who has logically privileged evidence for their occurrence? There can be no such person:

> "But when I imagine something, or even actually see objects, I have got something which my neighbour has not." - I understand you. You want to look about you and say: "At any rate only I have got THIS." - What are these words for? They serve no purpose. - Can one not add: "There is here no question of a 'seeing' - and therefore none of a 'having' - nor of a subject, nor therefore of 'I' either"? ...
> ... But what is the thing you are speaking of? It is true I said that I knew within myself what you meant. But that meant that I knew how one thinks to conceive this object, ... I think we can say: you are talking (if, for example, you are sitting in a room) of the 'visual room'. The 'visual room' is the one that has no

> owner. I can as little own it as I can walk about it, or look at it, or point to it. In as much as it cannot be anyone else's it is not mine either. In other words, it does not belong to me because I want to use the same form of expression about it as about the material room in which I sit. ((I, 398.))
>
> One might also say: Surely the owner of the visual room would have to be the same kind of thing as it is; but he is not to be found in it, and there is no outside. ((I, 399.))

We can now put the argument against private mental events in a superficially different form:

(i) Events which in principle no one could perceive, for whose occurrence no one could have any evidence, could not coherently be supposed to occur.

(ii) But we cannot make any sense of a person having evidence for the occurrence of his private mental events. So no one could have any evidence for the occurrence of such events - so they cannot coherently be supposed to occur.[41]

2.61 We may now consider a problem left by the Wittgensteinian line. (I mentioned a dilemma related to this problem in 2.43.) Assuming the PLA is successful, what are we to make of the intuition that we are introspectively aware of mental events occurring in us, and can report the occurrence of these in introspective reports? There is a widespread view that such events could not be perceived by others: and would therefore be in principle private to the person in whom they occur.

There are two aspects to this intuitive pressure against the PLA, expressible as these two views:

(i) A person can be aware of mental events occurring in him; and these mental events are reported by introspective reports.

(ii) These mental events (allowed by (i)) are accessible only to the person in whom they occur, and are in principle imperceptible to others.

It may seem odd that despite the logical force of Wittgenstein's

PLA, these two intuitive views are not dispelled by his argument. For these two intuitive views combine of course into the supposition that there are private mental events occurring in a person - they combine into a direct resistance to the PLA.

There is a reason why this resistance is not undermined by Wittgenstein's line. The style of the PLA is such that it does not need to engage closely with these intuitions - its logical force carries without requiring the detailed disarming of our intuitions. Moreover, what we can take as proven by the PLA is that we cannot have both (i) and (ii). The logical force of the PLA - at least, of the PLA as I have interpreted it - does not decide the issue of whether we are left with neither (i) nor (ii), or with (i) but not (ii). To decide this issue requires further argument. We either have to disarm the intuition that (i) obtains (leaving neither (i) nor (ii)); or disarm the intuition that (ii) must be a consequence of (i) (leaving (i) but not (ii)).

We can put this rather more simply and crudely: Wittgenstein's PLA does not explain to us: 'Look here: those events which you think are in principle private are only contingently private because ...'; or 'Those events aren't really occurring at all, because ...'.

Of course, many of wittgenstein's remarks can be taken as taking the line of denying (i) - or at least, of suggesting how we don't have to construe introspective reports as reporting the occurrence in us of mental events. But such a denial is not a direct logical consequence of the PLA. It only appears so if we feel that (ii) must be a consequence of (i). (Not an uncommon feeling, particularly among behaviourists: for clearly one of the main reasons for wanting to deny (i) is that allowing (i) seems to yield problematic consequences such as (ii).)

Behaviourist views take the form of denying (i). On such a view there are no events or inner processes of which we are introspectively

aware or which we report as occurring. In the context of Wittgenstein's remarks such a view can derive support from a modification of the Materialist Compatibility Point. Inner processes, whether contingently or necessarily imperceptible, would not be publicly observable. And provided it was maintained that such inner processes (even only contingently) were to remain behind the "shell" of public behaviour, there would be no need for them to feature in the interpretation of public meaning. Everything we say of each other and ourselves must derive significance from the observable "shell" of public behaviour, and would be logically compatible with there being no inner processes going on at all behind this "shell". So Wittgenstein's denial of (i) is perhaps not surprising given that his general picture of public meaning suggests that we must be able to interpret the meaning of everything we say as ultimately grounded on the "shell" of overt behaviour. (Ryle of course attempted to exhibit this grounding in more detail.)[42]

The trouble with denying (i) is that the resulting view is so counter-intuitive. Such a view would require that in a deep sense - going beyond our words (so to speak) - our mental life was just a "blank" - in the sense that we could not be said to be introspectively aware of any mental <u>events</u> occurring in us (and we would have to re-interpret introspective reports so that they can never be taken to assert the occurrence of such events). 'Train of thought', 'feeling of pain', 'after-imaging', 'visual imaging', and so on, could not figure in descriptions of particular dated events going on in people of which people can be introspectively aware. But there is a very strong intuitive view resisting such a supposition, and maintaining that we <u>can</u> be introspectively aware of events occurring in us described by such expressions.[43] There is a strong intuitive pressure for supposing that something genuinely happened while I was dreaming, which must have been an <u>event in me</u> of

which I can be retro-introspectively aware. And so on.

I am not too concerned here with increasing this intuitive pressure to embarrass the behaviourist line - for two reasons. In the first place, the view that (i) obtains is sufficiently well championed and brought to bear against behaviourism to need no further support here. Secondly, it seems fairly clear (as I mentioned above) that the main reason why the denial of (i) has been regarded as expedient in the face of this intuitive resistance, is that allowing (i) has been assumed inevitably to require (ii) as a consequence. In other words, in as much as a behaviourist line denying (i) might seem preferable to a line admitting (i), there may seem to be even more intuitive pressure for the entailment of (ii) by (i), than for the conservation of (i).[44] But if we can break the entailment of (ii) by (i), we no longer need shy away from admitting (i). I will attempt to break this entailment in Part 4; and then hopefully there will be no reason for wanting (i) put away.

This other option we have, of conserving (i) but without (ii), is the option which the mind-brain identity theory can be interpreted as taking. On the mind-brain identity theory (ii) is denied, it being maintained that mental events are in principle perceptible as neural events.

Clearly the problem for the mind-brain identity theory, taken as admitting (i) but not (ii), is to combat the intuitive pressure suggesting that (i) entails (ii). This particular intuitive pressure is not (at least explicitly) alleviated by Wittgenstein - indeed, we may suppose that because he felt it could not be alleviated, he chose the line (if he can be taken to have chosen <u>any</u> line) of denying (i). So it must be construed as the responsibility of the mind-brain view to <u>show</u> that admitting the mental events of (i) does not entail admitting (ii). Certainly, from the PLA we know that it would be wrong to admit (i)

and (ii). But this does not help us explain why the entailment appears to be justified, or why its justification is wrong: the PLA does not help (at least explicitly) to provide a perspective from which we can see what is wrong with the justification. Indeed, we don't even know from the PLA that the justification for the entailment is wrong: since (at least superficially) it is compatible with the PLA that the entailment does obtain, and that (therefore) we must deny (i).

We shall see in Part 3 that the accounts given in support of the mind-brain view either don't succeed in discharging this responsibility, or don't attempt to discharge it: the accounts available do not show that we need not construe the mental events allowed by (i) as private in principle - these accounts tend at best just to deny (ii). (As I have mentioned, I will try to discharge this responsibility in Parts 4 and 5.)

For the moment we should note the following general features of the context for the mind-brain view: -

(a) As was suggested in 2.43, if the PLA is right (and I am inclined to think that it is), we cannot construe the mind-brain identity theory in such a way that, if it is wrong, we then have ontological dualism. For if the PLA is right, we cannot coherently allow ontological dualism.

In other words, we cannot take the mind-brain view as trying to identify events which only may be physical with neural events. For if the PLA is right, we could not make sense either of events which are not physical, or of events which only may be physical. That's to say, the empirical nature of an identity hypothesis within the mind-brain identity theory, cannot be supposed to manifest itself through a view that only contingently is a certain event a physical event. Rather, the empirical nature of such an identity hypothesis must arise as it does for any

contingent physical event-identity hypothesis. (We shall see in Part 3 instances of this rather self-destructive confusion of supposing the empirical nature of mind-brain identity hypotheses to lie in the fact that mental events only may be physical.) It is therefore a problem for a view allowing (i) - indeed, the main problem - to explain how it is that we can construe the events admitted by (i), for which we don't appear (at least at present) to have physical descriptions, as nevertheless physical events. And of course this problem is essentially that of breaking the entailment between (i) and (ii).

(b) The mental events allowed by (i) would clearly be contingently imperceptible at present. We must therefore expect that we could only specify such mental events along the lines on which we tried to specify necessarily imperceptible mental events in Stage I. We should note that this way of extrinsically specifying events would be applicable to neural events (whether or not the mind-brain view is right): we may specify a certain neural event by using, for instance, 'N(v red)', analogously to 'P(v red)' (this is assuming that there are such typical neural events). We would in principle have available an intrinsic neural description of an event extrinsically specified in this manner. This indirect manner of specifying, applicable both to (supposedly) necessarily private mental events and contingently private neural events will be mentioned again in Part 3, where I discuss a relevant contention of Smart's regarding the "neutrality" of sensation reports.

NOTES TO PART 2

[1] Although these two aspects are often entwined in Wittgenstein's remarks, I will regard them as independent, since his view of meaning alone will be seen neither to disturb nor to be disturbed by the existence of private mental events.

[2] The distinction between the two stages resembles very roughly a distinction described by Strawson (in (1)) between a weak and strong thesis in Wittgenstein's Philosophical Investigations, corresponding to Stage I and Stage II respectively. But the correspondence is only rough: Strawson suggests for instance that the truth of the weaker thesis alone is sufficient to undermine an "other minds" sceptic (cf. Strawson (1), pp. 48 - 49). Whereas we shall see that both "other minds" and mind-body problems remain in Stage I.

[3] Quine, for instance, writes:

> In general, if a term is to be learned by induction from observed instances where it is applied, the instances have to resemble one another in two ways: they have to be enough alike from the learner's point of view, from occasion to occasion, to afford him a basis of similarity to generalize upon, and they have to be enough alike from simultaneous distinct points of view to enable the teacher and learner to share the appropriate occasions. A term restricted to squares normal to the line of sight would meet the first requirement only; a term applying to physical squares in all their scalene projections meets both. And it meets both in the same way, in that the points of view available to the learner from occasion to occasion are likewise the points of view available to teacher and learner on simultaneous occasions. Such is the way with terms for observable objects generally; and thus it is that objects are focal to reference and thought. ((Quine (2), p. 7.))

[4] Of course one may use a proper name when there is no referent of it - one may construe 'The Channel Tunnel' as a proper name, or, as has been done, invent a name, 'Chunnel', for the same purpose. But in these cases the use of the name is dominated by certain descriptions. The important point in this context is that the meaning of referring expressions need not be fixed by their referents, if any, nor do we in

all cases need access to their referents to know how we should use them.

[5] And this sort of point has often been made subsequently, notably by Wittgenstein, though he maintained in addition that there cannot be privacy of criteria at all; and it occurs in relevant contexts in, for instance, Ryle (p. 17), Strawson ((2) p. 99 and pp. 109 - 110) and in Hopkins (p. 140). Hopkins also mentions Frege's contribution in this area; but like Wittgenstein, and Dummett (whom I mention below), Hopkins does not entertain as reasonable the intermediate position of Frege's which is developed into Stage I, where the point in question is complied with while admitting the occurrence of private events.

[6] But the first point is the only one of the two which Dummett allows as a valid attack on an imagist or associationist theory of meaning (cf. Dummett (2), p. 638). Dummett denies, with Wittgenstein, that there can be any further intelligible distinction between "ideas" and "private ideas".

[7] This as we shall see is simplifying for the sake of smoothness in the initial projection of Stage I.

[8] I will use 'I' followed by the paragraph number to specify citations from Wittgenstein's Philosophical Investigations.

[9] In fairness to Stage I we have to regard these restrictive remarks of Wittgenstein's separately from those which I will regard as constituting the Private Language Argument. We can only try and minimise the degree to which this is unfair to Wittgenstein - for it is inevitable in a critical approach.

[10] To learn the use of 'red' might be to acquire the concept of redness; to learn the use of 'triangular' might be to acquire the concept of triangularity; but to learn the use of 'same' right is to learn that that is the word expressing a concept without the exercising of which we cannot conceive learning anything.

[11] We could, to make the model more analogous with the Stage I view, suppose that there is some "third board" - roughly analogous to the

"external world" - which is similarly connected to A's and B's boards. But for simplicity I leave this out.

[12] Though in Stage I we have to allow the possibility of mixing private and natural language sensibly, as in 'P'≠P"', we have to be careful. Take 'my sensation of red is vivid', 'my pain is getting worse'. If we construe these as containing publicly intelligible references to something private, we have to elucidate them in this manner: 'my sensation of red is vivid' would be quite infelicitous if it is intended as 'my P(v red) is vivid', for the same reason that it would be quite infelicitous to say of the chess model, 'Ap(knight) is white'. Rather, we have to say 'I have a P(v vivid red)', or use 'Ap(white knight)'. In other words, we have to segregate concepts of natural language which could not have sensible application to P's. Thus the private sensation of pain is infelicitously described as getting worse: rather, we have to say 'I have a private sensation of pain-getting-worse'. In the chess model we would have to describe a change of position in A's private system by, say, 'Ap(knight moves)', rather than 'Ap(knight) moves'. Certain concepts couched in natural language will not make sense applied to P's. But Stage I insists that though we have to be discriminating in this way, there is enough we can sensibly say about P's.

[13] This is not surprising in those taking Wittgenstein's line, since P's are banished altogether in this view. The restriction of isomorphism is not conspicuously required in the beetle analogy; nor is it emphasised in a card game example of Hopkins, who takes Wittgenstein's anti-Stage I line (cf. Hopkins, pp. 126 and 132). However, if we don't consider how the restriction of isomorphism is going to affect an account involving private events, then Stage I may be dismissed prematurely for wrong reasons.

[14] One reason may be that judgements of colour do not seem to be self-contained, in the sense that they seem to be related to other sorts of

judgement. It seems difficult to conceive of a colour-P inversion which would not affect any of my normal judgements about things other than colours - particularly difficult is the notion that P(v black) could be swapped for P(v white) without affecting any of my judgements. For it would seem possible perhaps to communicate that P(v black) is the "absence" of something, while P(v white) is the "presence" of something - "absence" and "presence" in this sense may be involved in a wide range of judgements (e.g., of smell, and so on) not just colour judgements. In which case someone with inverted colour-P's might be able to manifest this through his disagreement about P(v black) being the "absence" of something. I'm not sure though - perhaps not. At any rate, this is the sort of complication that may come in here. This sort of complication would certainly appear in connection with judgements of colour <u>intensity</u> rather than hue: since "intensity" applies with similar sense to judgements of weight, loudness, and so on. (So unless all these were inverted with respect to intensity, colour intensity inversion would be communicable.)

[15] Compare this with the following example: English written with the normal alphabet is isomorphic to English written by substituting for the n-th letter of the normal alphabet, the arabic numeral for n: 'bed' and '2,5,4' could be parts of two "systems" of English syntax. But they would not be as isomorphic as two systems of English syntax which are literally mirror images - we could not in this latter case separate the "content" which could be varied without altering the structure, from the structure.

[16] Thus there really seems no alternative in Stage I but to admit that unless I can imagine how my P(v red) could be other than it is, while still enabling me in principle to agree with others in all my colour judgements, I cannot then imagine how someone else's P(v red)

could be different from mine - notwithstanding the fact that I cannot compare my P(v red) directly with someone else's. It seems that I would be forced indirectly to suppose it was like mine.

[17] In case there is any discomfort associated with talk about images rather than imagings, we could easily transfer to talking about the event Y, and my private visual imaging P(v Y).

[18] Until the late fifties there was as a matter of empirical fact no known way of distinguishing between the north and south poles of a magnet without using 'right' and 'left' (as in whether a coil is wound clockwise or anti-clockwise) - indeed there was no known left/right asymmetric physical law. One such asymmetric law was discovered in 1956/7, by Wu (cf. Gardner pp. 223 - 237); but this law goes the opposite way in an anti-matter world. But since from the setting of the Ozma problem we can't tell whether the galaxy in question is made of matter or anti-matter (without presupposing the communication of 'right' and 'left'), there remains no known physical law by which we could communicate the meaning of 'left' (cf. Gardner p. 259).

[19] We could explicitly generalise Residual Privacy to cover perceptual events generally, but I will not do so here.

[20] We should note that Residual Privacy can only require a private distinction by a person if visual imagings are additional to neural events. In Part 4 we will see what happens to Residual Privacy when visual imagings are no longer construed as additional to neural events.

[21] E.g., cf. Sellars, pp. 14 - 18, Smart (1), pp. 59 - 61.

[22] The case of beliefs, desires, intentions and so on is rather different from the occurrence of private sensations, in that the former sort of mental entity (a) is not generally construable so that the mental entity is a private <u>event</u>, and (b) very plausibly does not require the supposition of the occurrence of a stream of private events in an

"organism" in order to be significantly ascribed to it. That's to say, it can very plausibly be maintained with Ryle that the ascription of this sort of mental entity only requires certain behavioural characteristics on the part of the "organism". But I shall have more to say about this later.

[23]This is reminiscent of some of Ryle's remarks, in that it would be a mistake to suppose, according to Ryle, that something <u>further</u> was required in order to attribute a mind to someone, when he showed this typical behaviour; this sort of mistake would be rather like that which:

> ... would be made by a child witnessing the march-past of a division, who, having had pointed out to him such and such battalions, batteries, squadrons, etc., asked when the division was going to appear. ... The march-past was not a parade of battalions, batteries, squadrons <u>and</u> a division; ((Ryle, p. 18.))

[24]This supposition is not coherent on Wittgenstein's view. I will mention this presently in 2.42. For the moment we may accept this uneasy and infelicitous supposition as a loose way of facilitating the presentation of his argument.

[25]Of course, with the Private Language Argument Wittgenstein tries to deliver a fatal thrust to this possibility: A would not even be able to ascribe the occurrence of private sensations to himself.

[26]Although Strawson (in (2)) emphasises that the manner of ascription may differ in the case of self-ascription from that in other-ascription without dividing the meaning of, say, 'pain' in 'A is in pain' and 'I am in pain', this can only be taken to vitiate the consequences of privacy, of Stage I, if a private process is <u>not</u> thereby being ascribed to either A or myself; and to establish this we seem to need something like the Private Language Argument.

[27] Dummett, in (1), overlooks this subtlety in Wittgenstein's position:

> For Wittgenstein, accepting the theorem is adopting a new rule of language, and hence our concepts cannot remain unchanged at

> the end of the proof. But we could have rejected the proof without doing any more violence to our concepts than is done by accepting it; in rejecting it we could have remained equally faithful to the concepts with which we started out. (((1), p. 430.))

[28] For instance by Boring, Turing, Putnam ((1), (2) and (3)), Davidson (2) and Hopkins. The examples of Davidson and Hopkins are most like mine, and Hopkins also uses his to illustrate a Wittgensteinian position.

[29] Clearly, in the view I am trying to illustrate there is no coherent distinction between an organism in which there occur private events and an organism in which there do not. I find it useful nevertheless to make this "incoherent" supposition to present a certain point. Providing we are aware of this difficulty the presentation will hopefully achieve its purpose. And given the nature of the point I want to make, there is no danger of this manoeuvre being question-begging.

[30] Wittgenstein's views on logical necessity, 'same', etc., are perhaps more easily understood in the context of Stage II. For in Stage II there is in a deep sense nothing "behind" the use of symbols by us. In a deep sense there is just behaviour. The idea of the necessity of '2+2=4' tends to come, according to Wittgenstein, from the idea that there is more to those symbols than any finite use of them - from some Platonic idea of the interpretation of '2+2=4'. Now, in what sense is it logically necessary that a machine programmed for arithmetical computation will output '2+2=4' and not '2+2=5'? In what sense is there a logical restriction that one of the organisms in the example above should have to think or decide that 2+2=4? All we could say perhaps is that the causal laws governing the outputs of these organisms are such that they use '2', '4', etc., as we use them. They could do otherwise, perhaps. (I'm not supporting this view, just indicating that it complements Stage II.)

[31] I will only in Part 5 consider more carefully the view that physicalism in the weaker form is sufficient for the stronger form. (Though I will recurrently hint at this sufficiency.) For the moment we may take this as no more than a suggestion.

[32] Arguments against this possibility may be of two types. First, it might be supposed that there are contingent scientific reasons why this is not possible. Secondly, it might be supposed that there are logical reasons why this would not be possible. Either case would of course obstruct the physicalist. (Though if the PLA is right we could not then take such obstructions to warrant ontological dualism.) There may not be the sharp division of types I suggest here, but at any rate I am assuming that there is not this sort of problem for physicalism.

[33] Thus Kenny seems to avoid the required commitment when he says: 'But if an experience is one that is kept secret, there is no reason to call an experience that is not kept secret a private experience.' (Kenny, p. 216). Cornman (in (2)) finds the engagement between the PLA and the mind-body issue rather different. For he regards the PLA as providing an objection to Smart's identity theory, by creating difficulties for the notion that sensation-terms can feature in genuine reports. (But this cannot follow directly from the PLA alone.) Yet although Cornman notes that the conclusion of the PLA is that 'No psychological expressions denote private entities' ((2), p. 120), he fails to observe how the PLA would affect the status of the mind-brain identity theory as an approach to ontological monism. For if the PLA is right, the identity theory can only be taken as a suggestion for the appropriate physical descriptions of mental events, in a context in which ontological monism is logically assured: not (as it is taken by Smart) a theory on which ontological monism depends.

[34] We shall later consider again this dilemma for those who are strongly inclined to follow (a), (b) and (c).

[35] Note that though Davidson's views suggest something like this, we shall see that Davidson must rely on something like Wittgenstein's PLA to complement his picture; for Davidson does not note the crucial role of privacy regarding the mental.

[36] Of course, I may prefer to decide that the colour of the patch has changed. But to decide this, backed by a genuine criterion for there having occurred a change in the colour, there would have to be some other genuine criterion in essence operating like the reference patch, to which I could then appeal.

[37] As we shall see later, Wittgenstein suggests that even my memory of the original private sensation is independent of it, so that I could be deceived by it. For the moment though we may accept that at least I can appeal to my memory, even though I cannot check whether my memory is right.

[38] It should also be clear that the characterisation of a private language is not just that it is used by one person: thus Rhees mistakenly defends the PLA on the basis that only languages used by more than one person are languages - cf. Rhees, p. 75.

[39] This sort of point is made by Kenny (p. 220).

[40] This sort of point is made by Ryle, of course (cf. p. 190ff.), and is discussed by Dennett, who describes the sort of mistake involved as positing a 'little man in the brain' (cf. Dennett, p. 99).

[41] It should be noted that the force of (ii) is not contained in the inherent claim that we cannot make sense of the two: a person and his private mental events. If this is taken as the force of (ii), an objection could be made as follows:

Consider a purely physical organism or structure M, rich enough not only to "monitor" and report events occurring in the world, but rich enough also to monitor and report the occurrence of some of the events occurring in it.

For M to perceive an external event E, it is absurd to suppose that M must perceive or somehow apprehend the event m occurring in it describable as 'M's perceiving E'. (This sort of point is made by Putnam ((1), p. 78) and Dennett (pp. 102 - 103).) Nevertheless, if M is rich enough, it may be able to monitor and report the occurrence of that event in M; and thus we may allow a sense in which M has evidence for the occurrence in it of m. But now we must suppose that there isn't a sense to saying that there is M and the events in it which it may have evidence for and report. And yet we allow a sense for saying that M has evidence for the occurrence of these events.

But here we don't have to construe the events in M as locked in relation to only one subject who can have evidence for their occurrence. All we need, to make sense of M having evidence for these events, is that we can suppose that there occur events in M which we could describe as "monitoring" other events in M. And of course, M's evidence isn't logically privileged (since any event in M must be conceived to occur logically independently of any other event in M). There is nothing mysterious about "M" even though there is no particular bit of the structure we could or would want to point to and say: 'That's the subject' (the referent of 'I' when M reports 'I ...'). We can coherently speak of "M" and "the events in M" even though these aren't logically independent - because we have a coherent sense for either on its own (as we can speak of either an object X persisting in time, or a sequence of events in X).

In the case in (ii), the private mental events have to be construed locked in (logical) relation to a subject who is aware of them. The force of (ii) is contained in the consideration that we can make sense of neither the "subject" nor "the events" - it's not that there aren't two, but that there isn't either: there is no other way of conceiving either "one" than as part of the incoherent "pair", so there can't be either "one" of this supposed "pair".

42 We should note, however, that it just does not follow that we would thereby be precluded from referring to contingently imperceptible events occurring in us with publicly grounded expressions of natural language (more or less in the Stage I manner). It only follows from the PLA that we could not refer to necessarily private entities, since we could not construe these coherently.

43 The case of 'visual imaging' (or other perceptual derivatives) is supposed to be less problematic for the denial of (i) than 'after-imaging' and some of the others, because the former is more readily construed as a grammatical hanger-on to descriptions of seeing in which a mental event or inner process occurring in a person doesn't appear to figure. Nevertheless, there is an intuitive pressure for supposing that something goes on in us when we see things - that 'something stops happening when I shut my eyes' asserts the cessation of a genuine happening which can't be the cessation of the event I was in the process of seeing, but must be the cessation of an event in me.

44 We should remember that (i) and (ii) requires ontological dualism, which is troublesome enough if it can be coherently entertained, without the result of the PLA (cf. 2.43) that it cannot be coherently entertained.

PART 3

THE MIND-BRAIN IDENTITY THEORY

The dualist has strong intuitive grounds for maintaining both (i) and (ii) of 2.61: for maintaining that people can be introspectively aware of mental events occurring in them (and can report these in introspective reports); and for maintaining that these mental events are in principle imperceptible, and so on.

I suggested in 2.61 that the mind-brain identity theory (which I will henceforth refer to as the Identity Theory) can only be acceptable as a monistic account which allows (i) (of 2.61) if it is shown that we can allow the mental events of (i) without having to suppose that they are in principle imperceptible. Indeed, in view of the incoherence of allowing (i) and (ii) (the incoherence suggested by the PLA), we should hope for a view which shows us that the dualist's arguments suggesting that (i) requires (ii) are incoherent - if, that is, the right monistic account is to be approached by allowing (i).

But we shall see that this is not achieved by proponents of the Identity Theory. I will exhibit this failure not by means of an exhaustive survey of arguments for the Identity Theory, but by tracing the general strategy of discussions in this area. (There may be arguments that I don't know of which explicitly fulfil the requirements I set for the Identity Theory. Naturally I have to assume there are none.)

In 3.1 I will mention the weakness in the Identity Theorist's general strategy, and in 3.2 to 3.7 I will indicate how this weakness manifests itself in some detailed confrontations between the Identity Theorist and his objectors.

Because I will frequently refer to (i) and (ii) of 2.61, I will henceforth refer to these as (o) and (p) respectively.

Also to simplify matters I will use 'ψ' (psi) and 'ϕ' (phi) rather

loosely: ψ is either a particular (dated) mental event or a certain type of mental event (undated); ϕ is a neural event (dated or undated - naturally in '$\psi = \phi$' we take it that 'ψ' and 'ϕ' match). I will also loosely speak of 'ψ' as connoting: really I am speaking of the connotation of the appropriate mental description for which 'ψ' is a dummy - but it will simplify if I stick to 'connotation of "ψ"', and likewise 'connotation of "ϕ"'.

3.1 Problems of the General Strategy

It is possible to discern the following strategy in the original conception of the Identity Theory as an approach to physicalist monism (e.g. in Place, Feigl ((1) and (2)), and Smart (1)).

The Identity Theorist agrees with the dualist that mental events may logically possibly be additional to neural events, and physical events generally - as the heavenly body observable just before dawn in the eastern sky once seemed to be additional to the heavenly body observable just after sunset in the western sky. And then, on analogy with the case of Venus, and certain scientific identity hypotheses, the Identity Theorist suggests that those events which the dualist supposes are additional to physical events are in fact neural events.

3.11 There are two aspects of this sort of view which are very troublesome as regards the internal stability of the Identity Theory.

(i) The analogy with Venus and with scientific hypotheses seems to suggest that the Identity Theorist is prepared to accept that 'ψ' has the same connotation for him as for the dualist - the difference between them being as regards how they take the reference of 'ψ'.[1]

Thus it was hoped that on analogy with scientific hypotheses, the case of Venus, etc., the mere fact that the connotation of 'ψ' is different from that of 'ϕ' would prove no logical obstacle to the identity hypothesis that $\psi = \phi$. (I will henceforth refer to such a hypothesis ('$\psi = \phi$') as an (or the) Identity Hypothesis.) It was hoped that just as the connotation of

'the heavenly body observable just before dawn in the eastern sky' may be shared by both someone who does and someone who does not believe that it refers to the heavenly body observable just after sunset in the western sky, a similar sharing of connotation would allow the Identity Theorist to maintain that those events which the dualist supposes are additional to neural events are in fact neural events.

And now the problem is this. If the connotation is not shared, the Identity Theorist is in danger of being accused of failing to establish ontological monism, since he is not identifying <u>those events</u> taken (by the dualist) to be additional to physical events, with physical events. Although he is making <u>an</u> identification, it would be retorted that there are still other events, not referred to by either side of the Identity Hypothesis, which are additional to physical events. On the other hand, if the Identity Theorist shares the dualist's connotation, the problem is then that the dualist will argue that given his connotation, mental events <u>could not</u> be neural events (or physical events generally). This is in contrast to the case of Venus, where it doesn't appear that the heavenly body observable just before dawn in the eastern sky <u>could not</u> be the heavenly body observable just after sunset in the western sky.

(ii) The Identity Theorist allows that dualism is logically possibly true. He allows that mental events logically possibly may be additional to neural events and physical events generally. Thus he maintains that only as a <u>matter of fact</u> are mental events physical events. (Actually there is a persistent failure in the original conception of the Identity Theory to distinguish carefully between 'mental events are as a matter of fact physical events' and 'mental events are as a matter of fact neural events'. I will mention this in a moment. But clearly, if dualism is possible, if mental events possibly may be additional to physical events, then only as a matter of fact are mental events physical events.)

Thus Smart says that the issue between the brain process theory and

dualism seems to be like the issue between certain conventional palaeontological theories of the origin of the earth, and Gosse's theory: a theory maintaining that the earth was created about 4000 years B.C., but in exactly the form it would have been like in 4000 B.C. if it had been in existence before then and undergoing conventionally accepted evolution (cf. Smart (1), p. 66). So Gosse's theory and conventional palaeontological theory are both compatible with the earth's being as it is now. Smart says:

> If it be agreed that there are no cogent philosophical arguments which force us into accepting dualism, and if the brain-process theory and dualism are equally consistent with the facts, then the principles of parsimony and simplicity seem to me to decide overwhelmingly in favour of the brain process theory. (((1), p. 66.))

This suggests (a) that empirical evidence is in principle compatible with both the Identity Theory and dualism; and (b) that dualism is logically possible: that mental events logically possibly may be additional to physical events.

If empirical evidence may be compatible with both dualism and the "matter of fact" maintained by the Identity Theory (a rather odd "matter of fact"), empirical evidence cannot decide the dualism/Identity Theory issue - only logical considerations can (assuming that empirical evidence turns out to be compatible with the Identity Theory). And yet the Identity Theory apparently allows that dualism is logically possible: when, as it were, logical space is really not large enough to hold them both. A scientific factual theory can afford to allow an (empirically) incompatible rival: facts can decide the issue between them. But a theory which cannot be favoured with respect to its rival on factual grounds, yet is logically incompatible with its rival, must confront its rival on logical grounds. Yet the only logical ground Smart suggests for favouring the Identity Theory is parsimony; but it just won't do to suppose that parsimony should decide between two <u>logically incompatible</u> views: for instance, parsimony shouldn't decide the question of whether or not we are to admit transfinite numbers.

It may seem very chivalrous of Smart to allow the rival view as

logically possible (certainly his rival doesn't reciprocate). But I think Smart's motivation for this is a feeling (based on a confusion) that he must allow that dualism is logically possible.

It looks as though Smart supposes that if it is logically necessary that mental events are physical events, then he would have to claim that a mental description (figuring in his Identity Hypothesis) is meaning-reducible to the physical description (figuring in the Identity Hypothesis). He feels, perhaps, that because he wants to maintain an Identity Hypothesis '$\psi = \phi$' which is contingent, then it must be logically possible that ψ isn't ϕ, in which case there would be two events, one mental and one physical; so dualism must be logically possible.

But it does not follow from the contention that mental events cannot possibly be additional to physical events, that an Identity Hypothesis '$\psi = \phi$' does not assert a matter of empirical fact. It doesn't follow in the case that we can exhibit ψ as a physical event, which may as a matter of fact accept the neural description 'ϕ'. (This is after all what happens when 'a' and 'b' are physical event descriptions, and we say that contingently a = b.) In other words, the supposition that mental events in human beings are as a matter of fact neural events does not need to carry along with it the supposition that mental events in human beings are (only) as a matter of fact physical events. But the Identity Theorist tends to lump both these suppositions together.

There is another reason why Identity Theorists may want to regard dualism as logically possible. It is a result of this sort of thought-experiment: one imagines that it turns out that one's skull is empty, indeed, that one has no central nervous system, while the rest of the body is normal, and one's experiences are perfectly normal. So therefore ontological dualism is logically possible.[2]

Regarding the conclusion here, of the possibility of dualism, and supposing for the moment that the thought-experiment forced us to it, we

can make this comment: in the case of a person who has introspective evidence of his mental events, but has no central nervous system, the mental events are in effect disembodied - the person still has a body, but the effect of this case is precisely that of a case where we explicitly take a completely disembodied person. Well, the subject in either the partially or the totally disembodied person would be logically tied to his mental events: he would in principle be the only subject in the universe who could have evidence (in the partial disembodiment, direct evidence) of the occurrence of his mental events. This possibility is ruled out by the PLA as not really coherent.

As regards the thought-experiment that tried to force us to allow the possibility of dualism, we can say this: there is a supposition which if true renders the thought-experiment confused. The supposition is that mental events are neural events. For if mental events <u>are</u> neural events, we can imagine that our skulls are empty, but if our skulls were empty we couldn't imagine.

It is logically possible that a phenomenon might occur without some other phenomenon. And it is logically possible that a phenomenon which in fact occurs should not have occurred at all. But we cannot allow as logically possible that a certain phenomenon occurs though <u>it</u> doesn't occur.

If consciousness <u>is</u> a cerebral process (and let's put aside here any intuitive misgivings about this), then we cannot allow the possibility that I should have been introspectively as I in fact am, though cerebral processes didn't occur: for to be introspectively as I am, on the brain process theory, is for my cerebral processes to be as they are - there are not <u>two</u> phenomena (<u>ex hypothesi</u>), but one phenomenon. So for me to imagine being introspectively as I am, but with an empty skull, would (<u>ex hypothesi</u>) be to imagine that cerebral processes occur in me but that my skull is empty. (We are assuming the brain is nowhere else.) If there are not two phenomena, the stream of consciousness and the neural stream, then the possibility of a person being introspectively as he is but with no brain is simply the possibility of a

physical structure being as it is but not being as it is: an impossibility. We could not then have been different in respect of neural features without being different in respect of introspectively given features. Of course, if there are two phenomena, if ontological dualism is true, then the possibility in the thought experiment is a real possibility. But it is not a real possibility independently of whether ontological monism is true or not; so the thought-experiment does not force us to accept, against Wittgenstein, that ontological dualism is logically possible: for if dualism is logically impossible and ontological monism is true, then the thought experiment does not mention a real possibility.

We should note two points though. First, we block the thought-experiment on the supposition that there is just one phenomenon, the neural stream. Really, the intuitive unease at the way the thought-experiment is blocked is based on a view which cannot accept what is introspectively given and the neural stream as one phenomenon. This unease is not in the slightest bit relieved by the simple consideration that it must be wrong (that the PLA is right) - to relieve this unease we must take it apart.

Secondly, even supposing that ontological monism is true, and that if we are genuinely aware (or able to be aware) of mental events occurring in us, these events must be physical, it is still logically possible that we should have made different discoveries from the ones we in fact have made. It can still be regarded as an empirical discovery that the physical descriptions of mental events are neural descriptions. Thus suppose that someone had never heard of the brain, and had no knowledge of anatomy, but formulated and accepted the PLA. Then he might be able to deduce that the mental events he was introspectively aware of, not being logically tied to a mental subject of awareness in him, were available for direct inspection (in principle) by other people, and so on - that there must be physical descriptions for them and that they could not be disembodied. He might be able to deduce that his mental events are physical events occurring somewhere (probably

physically inside his body), even if he could not imagine what physical descriptions they would have, what they would be like when perceived. (He might, alternatively, abandon the view that he is genuinely aware of any events occurring inside him which he can't (yet) physically describe - abandon (o).)[3]

3.12 At any rate, Smart (and others) accept that dualism is logically possible. And the general idea then is that the Identity Theorist doesn't have to undermine dualism.

What the Identity Theorist should do to engage satisfactorily with the dualist, while avoiding having to suppose either that mental events could not be physical, or that they may not be physical, is show that those events which seem to the dualist to be additional to physical events seem so on mistaken grounds: by showing the incoherence of the justification for taking (p) as required by (o). The Identity Theorist would thereby be showing that mental events allowed by (o) can and must accept physical descriptions: they are and must be physical events.[4] Nevertheless, such a view can retain the feature that ascertaining what physical descriptions they in fact have is an empirical matter. This strategy would avoid (a) having to share a connotation for '$\not\phi$' in which '$\not\phi = \phi$' is quite implausible; (b) the supposition that the Identity Theorist doesn't have to show that dualism is incoherent - a supposition made by some Identity Theorists hoping that the parsimonious nature of their view outweighs the dualist's protestations, or at least, hoping that the dualist cannot prove them wrong; (c) being accused of not identifying the potentially or actually troublesome events (the events taken by the dualist as additional to physical events), with physical events.

We will find that Identity Theorists only half-heartedly, if at all, undertake to show that (o) doesn't require (p). And that as a result, either explicitly or implicitly, they are interpretable as taking one of the three undesirable alternatives outlined just now.

For what seems to happen in the Identity Theory controversy is this. The Identity Theorist starts off hoping to engage with the dualist - essentially by allowing (o).

Then the objector points out that the mental events of (o) are necessarily imperceptible, and so on: he maintains that (p).

(We know from Part 2 that to suppose that there are necessarily imperceptible mental events is incoherent - but we cannot yet say whether this means that we can allow neither (o) nor (p), or whether we can allow (o) and not (p). I will find it convenient therefore to put in abeyance for the moment the fact that we must must not allow (o) and (p).I will thus treat the objector's arguments as though it were coherent to allow private mental events. This is also convenient because in general the Identity Theory controversy is to some extent naive to the full impact of the PLA.

So I will here in Part 3 accept that there are intuitive pressures for maintaining both (o) and (p), and generally suspend the view that it's incoherent to maintain both - thus allowing for the moment, with some Identity Theorists, that ontological dualism is coherent. In Parts 4 and 5 I will try to tidy this up by exposing as confused the objections to the Identity Theory based on the supposed requirement of (p) by (o).)

Now, the Identity Theorist seems to be rather at a loss when it comes to showing that (o) doesn't require (p). But he recovers his poise by exploiting various technicalities relating to the objections presented so as to cast doubt on their efficacy against the Identity Theory.

I will argue that this sort of maneouvre, in the absence of arguments against (p) being required by (o), leaves the Identity Theorist vulnerable to the suggestion that he can no longer be maintaining (o).

Indeed, I will suggest that there are discernible in the literature two forms of the Identity Theory (neither of them being the form I suggest as most attractive). The first form may be taken as admitting (o), but also as admitting that dualism is possible. This I have suggested is an internally

unstable view - quite apart from its vulnerability to dualist objections. And the second form, though an internally stable view, does not allow (o), and does not attempt to engage with, or does not successfully dispel, the intuitive pressures for maintaining (o). This second form, though immune from objections based on (p) (since it does not maintain (o)), is not as I say, immune from the pressure for maintaining (o) (which dualist objections can be taken as at least implicitly displaying). More importantly, the second form can only be interpreted as differing from Wittgenstein's position in respect of a meaning-analysis of mental reports.

Although there is evidence that Smart wanted to keep (o), I will suggest that he seems most readily interpretable as taking a line of the second sort, in which (o) is abandoned.

Generally what I want to say regarding the objections and counter objections I deal with can be said by taking an event Identity Hypothesis containing the loose 'ϕ' and 'ϕ'. Where the form of the Identity Hypothesis assumes more importance I will adapt accordingly. But the majority of objections to the Identity Theory which I will consider apply to Identity Hypotheses containing descriptions of particular dated events: and the success of this sort of Identity Hypothesis is both necessary and sufficient for ontological monism.[5]

3.2 The Identity Hypothesis and Leibniz's Principle

In formulating the Identity Theory, Place was attracted by the form of certain model scientific hypotheses. For instance, the hypothesis that (a) there are not, as a matter of fact, two things, one a cloud, one a mass of tiny water particles, but just one thing; (b) this one thing is a mass of tiny water particles and nothing more. Smart was also attracted by this sort of example, though he was more sensitive to the possibility of a disanalogy between the mind-brain identity theory and the cloud-water particle identity theory - in particular, the possibility that there might be logical obstacles to the supposedly analogous contention that the stream

of consciousness is nothing over and above a cerebral process.

Suppose x is W if it is composed of water particles; and x is C if it is visible to the naked eye as floating in the sky. Clouds turn out to be both C and W. The analogy suggests that if the Identity Hypothesis is true, certain events would be both mental and neural. For there is no elimination of the properties we think x and y to have when it turns out that x = y. (Except, perhaps, for the "property" we may have thought x and y to have: "of not being one and the same".) There is only a reduction in the number of things we thought there to be.

This sort of point is made by Stevenson - essentially he emphasises the symmetry involved in the Identity Hypothesis, as revealed by Leibniz's Principle (cf. Stevenson, pp. 88 - 89). Stevenson suggests that the supposedly problematic properties of mental events would not be eliminated just on the supposition that mental events are neural events.

There are two sorts of objection inherent in Stevenson's point. One is an objection to the Identity Theory as an approach to the strong form of physicalism: the problematic properties of mental events may not actually obstruct the Identity Hypothesis, but they may lie problematically outside the physicalist scheme. The second objection inherent in Stevenson's suggestions (though as an explicit formulation of this objection his suggestions would be inaccurate) is an objection to the Identity Theory as an approach to ontological monism (and physicalism in its weaker form). If the properties of mental events were a problem in this respect, it's not so much that they would not be eliminated through the Identity Hypothesis - they would obstruct the Identity Hypothesis.

I will only consider here in Part 3 objections to the Identity Theory as an approach to ontological monism - this is the lesser claim of the Identity Theory, so if it is deficient in this respect the Identity Theory will also be deficient regarding the stronger claim.

The possibility of the second sort of objection, obstructing the

Identity Hypothesis, seems to have been overlooked by Place, who thought originally that the 'thesis that consciousness is a process in the brain is put forward as a reasonable scientific hypothesis, not to be dismissed on logical grounds alone.' (Place, p. 42). Thus he says:

> Those who contend that the statement 'Consciousness is a brain process' is logically untenable, base their claim, I suspect, on the mistaken assumption that if the meanings of two statements or expressions are quite unconnected, they cannot both provide an adequate characterisation of the same object or state of affairs. ((Place, p. 45.))

But this of course is much too optimistic - there are logical grounds on which the Identity Hypothesis might be denied or qualified, particularly those invoking Leibniz's Principle, which I will take as:

$(x)(y)((x = y) \supset (F)(Fx \equiv Fy))$.

Thus if Fx.Gy, it might still be possible that x = y. But it's possible that x = y only if it's possible that Fx.Gx - if one thing could not have the properties F and G because these are in some way incompatible, as where $Fx \equiv \sim Gx$, then if Fx.Gy it would not be possible that x = y.

Consequently we have room for the general objection to the Identity Theory, that the properties which $\not\alpha$ and $\not\phi$ are known or supposed at present to have render it impossible that they could be one and the same.

Thus the strategy of the objector is then to find at least one F such that $F\not\alpha . \sim F\not\phi$. And the Identity Theorist's defence is then to consider the F's presented, and in each case deny either that $F\not\alpha . \sim F\not\phi$, or that $(Fx . \sim Fy) \supset (x \neq y)$. In other words, the Identity Theorist tends to argue either against the truth of '$F\not\alpha$' or '$\sim F\not\phi$', or against the applicability of the predicate F in Leibniz's Principle.

Before considering the properties of mental events used against the Identity Hypothesis, I will discuss an attempt by Smart to pre-empt such objections. But we shall see that this attempt already shows the inherent instability of his view, which seems to straddle two rather different forms of the Identity Theory.

3.3 Smart's Analysis of Sensation Reports

Smart's response to Stevenson's objection was to suggest, in effect, that the defining properties of $\not{\phi}$ - roughly, those properties $\not{\phi}$ is taken to have whether or not $\not{\phi} = \phi$ - need not and should not be considered as incompatible with $\not{\phi}$ being a neural event (nor need these defining properties lie outside the physicalist scheme).

(This sort of pre-emptive move could take the form of arguing that (o) does not require (p): that the mental events we can be introspectively aware of are not in principle imperceptible, and that we can conceive that there may be physical descriptions of them. This is what we saw the Identity Theorist had to do if his theory is to underpin physicalism. Smart's proposal may perhaps be interpreted as taking this form - but it is not successful.)

Smart's support for his reply to Stevenson is derived from an attempt to show that introspective reports and reports of mental occurrences - sensation reports in particular - mention no properties or entities which are explicitly incompatible with the Identity Hypothesis. At the same time, such reports are in no danger of being synonymous with or meaning-reducible to reports of neural events involving neurophysiological descriptions.

In his early paper Smart anticipates Stevenson's objection, and outlines the basis of his later reply:

> Now how do I get over the objection that a sensation can be identified with a brain process only if it has some phenomenal property, not possessed by brain processes, whereby half of the identification may be, so to speak, pinned down? ... My suggestion is as follows. When a person says, 'I see a yellowish-orange after-image', he is saying something like this: 'There is something going on which is like what is going on when I have my eyes open, am awake, and there is an orange illuminated in good light in front of me, that is, when I really see an orange.' ((Smart (1), p. 60.))

'I see a yellowish-orange after-image' is taken as a report of a mental occurrence. Now, if we spell out exactly what occurs when someone has a yellowish-orange after-image, there should, according to the objector to the Identity Theory, be mention of some feature of the reported occurrence

rendering it impossible that $\not\phi = \phi$.

But Smart finds that his elucidation of the visual sensation report is quite compatible with a neural event being reported. There is mention of a stimulus condition, and of something occurring, but the elucidation statement could be made true by the "something" being a neural event. (And Smart's elucidation is of course generalisable to any sensation report.)

Smart suggests that these considerations stress and to some extent explain the fact that an examination of what it is to be a sensation reveals sensations as curiously and elusively neutral (cf. Smart (1), p. 61); he suggests that the only properties we are committed to attributing to them are those relating to the stimulus conditions by which they are specified.

This line is pursued in Smart (2) (p.93) by developing an analogy introduced in the earlier paper ((1),p.60). The analogy is between the neutrality of sensation reports and the neutrality of the report 'someone telephoned' - the latter not being synonymous with 'the doctor telephoned' even though the "someone" was the doctor. Smart suggests that a sensation report is explicitly neither a neurophysiological report nor a non-neurophysiological report, in the way that 'someone telephoned' is neither medical nor non-medical: even though 'someone telephoned' is made true by the doctor telephoning. For on Smart's elucidation, a sensation report can be made true by the occurrence of a brain process.

I find this proposal of Smart's inadequate, however, for the following reasons.[6] Smart's elucidation along the lines 'There is something going on which is like ...' might be a sensation report, just as 'someone telephoned' might be a "medical" report. But Smart's elucidation statements should be good enough to be recognised as elucidations of sensation reports. Indeed, Smart admits: 'Stevenson will ask me how then I distinguish the class of sensations. I reply that they are the things, whatever they are, that are reported in a certain recognisable class of utterances.' ((2), p. 94).

It will not suffice for an elucidation of a sensation report, an elucidation of the form '$(\exists x)(Ex)$', to be true if x is the appropriate

sensation. The sort of elucidation required must be a meaning-analysis. So at least we should have that $(\exists x)(Ex)$ iff I have a yellowish-orange after-image. But this bi-conditional cannot be achieved without significantly departing from the neutrality of Smart's elucidation.

Suppose I reported 'There is something going on (in me) which is like what goes on when I have my eyes open, ... and I see blood'. Is this a visual sensation report, or a report of nausea, or a report of an event in my eyes? When I say that I have a visual sensation of seeing blood, I mean to and do rule out the possibility that my report is of the occurrence of a feeling of nausea (which I typically have when I see blood, say). And to remove the lack of specificity that has been introduced into the 'There is something ...' sentence, I have to rule out various things the "something" could be: as when I insert 'visually' to qualify how the something is like what goes on in me when I see blood - ruling out the feeling of nausea as being reported. We just could not make the elucidation recognisable as an elucidation of a sensation report without ruling out the replacement of 'something' (or rather, an appropriate definite description in an adaptation of the quantified sentence) by 'ϕ' salva veritate, even though $\not\phi = \phi$ - for, as Hinton puts it, the elucidation sentence would become non-extensional, containing something like 'I find like' instead of 'like' (cf. Hinton, p. 252).[7]

To remove the lack of specificity involves a departure from neutrality. Sensation reports are just not elusive enough to be unrecognisable as such, nor do they leave things as open as on Smart's elucidation.

3.31 There is a suppression involved in Smart's proposal, which takes the form of overlooking the following distinction: between (a) specifying an X(-sort-of-thing) as a particular X, say X', with the tacit understanding that it is an X (a report which does this is interesting in as much as it tells us which X it is, rather than what it is to be an X); and (b) specifying what it is to be an X. To specify a particular (having of a) sensation one has to resort to a specification of stimulus conditions; but to specify

something as a (having of a) sensation, rather than some other process, then, even though a (having of a) sensation might turn out to be this "other" process, something more general and explanatory should be said. But Smart's elucidation succeeds in (a) only - whereas it ought to succeed in both (a) and (b). And clearly there is going to be an objector to the Identity Theory who will want an elucidation of a sensation report to read something like 'There is something going on (in me), which is in principle private to me, imperceptible to others ... and which ...'. Smart is able to avoid an explicit rejection of this sort of elucidation, but only at the cost of failing to elucidate sensation reports. An elucidation involving privacy would be incompatible with the supposition that neural events are reported in introspective or sensation reports.

Certainly, as I noted in 2.61, we could specify contingently imperceptible neural events in the manner 'N(v red)': i.e. in terms of extrinsic stimulus conditions and so on (supposing there are such typical neural events). And we can consider a truly neutral specification 'E(v red)': 'The type of event occurring ...'. But the dualist, in particular the psycho-physical parallelist, will suppose that there are two radically different sorts of typically occurring events in a person: N(v red) and P(v red); and that we are introspectively aware only of the latter sort of event.

An explicit rejection of a privacy elucidation could be achieved if it was shown that the dualist is mistaken in thinking that the events we can be introspectively aware of and report in introspective reports must be (in principle) private. If, in other words, it was shown that (o) doesn't require (p). But Smart does not achieve this.

Smart has to deny implicitly that (p) obtains. But in so doing he seems forced to deny that we can be introspectively aware of the mental events which we report as occurring, and which are supposed to be neural events.[8] For if we were introspectively aware of neural events, we would be introspectively aware of events with neurophysiological properties. And,

unable to surmount the difficulty here, Smart seems forced to take the line that we are not really aware of any intrinsic qualities of the events occurring in us:

> Raw feels, in my view, are colourless for the very same reason that something is colourless. This does not mean that sensations do not have plenty of properties, for if they are brain processes they certainly have lots of neurological properties. It only means that in speaking of them as being like or unlike one another we need not know or mention these properties. ((Smart (1), p. 61.))

And Smart says of his defence by the neutral elucidation of sensation reports, that its strength 'depends on the possibility of our being able to report that one thing is like another without being able to state the respect in which it is like.'. But what exactly are we introspectively aware of then?

The most plausible way to interpret Smart's theory now is to suppose that he denies (p) and denies (o); but that he is maintaining that instead of construing expressions which purport to refer to mental events or inner processes as not being genuine referring expressions - instead of taking a Wittgensteinian "avowal theory" of sensation statements, or some such view - we should construe such expressions as genuine referring expressions : referring to neural events. This interpretation of the view to which Smart is inclined in the face of Stevenson's objection is confirmed by the statement I cited earlier:

> Stevenson will ask me how then I distinguish the class of sensations. I reply that they are the things, whatever they are, that are reported in a certain recognisable class of utterances. (((2), p. 94.))

Which conspicuously avoids commitment to the view that we are introspectively aware of sensations.

3.32 A similar difficulty affects Place:

> ... statements about pains and twinges, about how things look, sound, and feel, about things dreamed of or pictured in the mind's eye, are statements referring to events and processes which are in some sense private or internal to the individual of whom they are predicated. The question I wish to raise is whether in making this assumption we are inevitably committed to a dualist position in which sensations and mental images form a separate category of processes over and above the physiological processes with which they are known to be correlated.

> I shall argue that an acceptance of inner processes does not entail dualism... ((Place, p. 43.))

But Place's treatment of the problem of introspection and introspective reports is rather like Smart's: in the end Place has to rely on the fact that the events which the dualist supposes are non-material can only be specified <u>in terms of</u> material things (namely his environment):

> ... when we describe the after-image as green, we are not saying that there is something, the after-iamge, which is green; we are saying that we are having the sort of experience which we normally have when, and which we have learned to describe as, looking at a green patch of light. ((Place, p.50.))

Place argues that:

> ... the problem of providing a physiological explanation of introspective observations is made to seem more difficult than it really is by the 'phenomenological fallacy', the mistaken idea that descriptions of the appearances of things are descriptions of the actual state of affairs in a mysterious internal environment. ((Place, p. 42.))
>
> This logical mistake, which I shall refer to as the 'phenomenological fallacy', is the mistake of supposing that when the subject describes his experience, when he describes how things look, sound, smell, taste,or feel to him, he is describing the literal properties of objects and events on a peculiar sort of internal cinema or television screen, usually referred to in the modern psychological literature as the 'phenomenal field'. If we assume, for example, that when a subject reports a green after-image he is asserting the occurrence inside himself of an object which is literally green, it is clear that we have on our hands an entity for which there is no place in the world of physics. ((Ibid., pp. 49 - 50.))

We can take the "phenomenological fallacy" to be the sort of mistake exposed by Wittgenstein's PLA. (Though we may note that even a Stage I view would not maintain that we can or do describe in public natural language the intrinsic qualities of private mental events.)

But then we have the problem, unsolved by Place and Smart, that it is not clear that we <u>can</u> be introspectively <u>aware of</u> events occurring <u>in us</u>. And it is especially difficult to suppose that we can be introspectively aware of the mental events which (supposedly) are neural events. For if I can be introspectively aware of some event occurring in me when it is said, or I say, that an after-imaging occurs in me, then the fact that we have to specify that event occurring in me <u>in terms of</u> external stimulus conditions does not make it any easier for me to conceive that the changes I am

introspectively aware of are neurophysiological changes: - are those processes which are in principle perceptible to other people if they observed my brain.

It seems therefore that we are forced to construe Place's view, and the view Smart appears to take in response to Stevenson's objection, as views in which (o) is denied. There is a subtle shift from: 'Those events of which we can be introspectively aware and which seem to the dualist to be additional to physical events are in fact neural events' to: 'Those expressions which are taken by the dualist to refer to non-physical events (and by the behaviourist not to refer at all) are to be construed as referring to neural events'.

3.4 Two Identity Theories

The ambivalence we saw in Smart's position, and which also occurs in Place's position, can be usefully simplified as a conflict between the following (crudely depicted) positions:

I. There are mental events we can be introspectively aware of: those mental events which the dualist supposes are additional to physical events, being in principle private and imperceptible etc... Indeed, they may be additional to physical events, and in principle private etc. - we do not dispute this. However, we maintain that as a matter of fact they are physical events: neural events. (So in fact they are in principle perceptible.)

II. (a) There are no mental events which are in principle private and imperceptible etc. ... Nor are there any mental events we can be introspectively aware of. (Rather, we should more felicitously say: sentences such as 'I am introspectively aware of...' should be construed as grammatical derivatives of 'there occurs in me...'. (Cf. 2.42 for a similar problem of infelicity in the making of a similar sort of point.))

(b) Expressions which we may be tempted to construe as referring to the events denied in (a) are to be taken as referring to neural events.

The trouble with the I-Identity Theory is essentially this: such a view fails to cope with the dualist's view that mental events are necessarily imperceptible: that (o) and (p). The mere hypothesis that mental events are neural events, though certainly denying (p), cannot by itself be taken as an argument against the view that (o) requires (p).

For the I-Identity Theorist allows that possibly (o) and (p); and so allows to persist an intuitive view which supposes that the Identity Theory is quite implausible. The I-Identity Theorist has to rely on the possibility that the objector may not be able explicitly to prove that (o) requires (p).

It may be supposed that the onus is on the dualist to prove that (o) requires (p). Certainly: in so far as concerns proving the Identity Theory to be false. But the onus is on the I-Identity Theorist to prove that the dualist view is false: the onus is on the I-Identity Theory to solve a problem of mind - if it is intended to solve a problem of mind.

In contrast with a view which accepts that neither (p) nor (o) obtains, the I-Identity Theory is presumably intended to engage with and solve a certain ontological problem: presented by the intuitive pressure to maintain (o) and (p). But the I-Identity Theorist can hardly be commended for undertaking this solution by demanding that it be proved that (o) requires (p). For this would be tantamount to suggesting that it is not proven by those maintaining that there is a problem (a problem implicitly recognised by the I-Identity Theory), that their problem has a sound basis.

The II-Identity Theory, on the other hand, denies both (p) and (o). It is an interesting variant of Wittgenstein's view, providing as it does an alternative meaning-analysis of introspective reports. I don't here want to criticise this sort of view directly. But it should be noted that such a view does not dispel the intuitive reluctance to part with (o). (As noted

in 2.61, the main reason for denying (o) on such a view comes from the difficulty of allowing (o) without (p).) And, as with Wittgenstein's view, the II-Identity Theory smacks of "blank automatism": it can most readily be taken as a linguistic rather than "metaphysical" alternative to behaviourism.

So both the I- and II-Identity Theories allow what may be considered as the crucial problem of mind to remain: the problem of how to keep (o) without (p).[9]

It seems clear that Smart started out as a I-Identity Theorist.[10] But unable to manage the difficulties involved in allowing (o), Smart seems forced into a II-Identity Theory.

I will further illustrate the difference between the two Identity Theories in 3.7. In 3.5, 3.6 and 3.7 we will confirm the anticipated difficulties for the Identity Theory. These difficulties manifest themselves in confrontations between the Identity Theory and the standard objections to it. These objections arise, as we might expect, from a view maintaining (o) and (p).

Thus it is argued that a person has logically privileged access to his mental events, which are in principle imperceptible to others, and (concomitantly) cannot be located spatially or generally partake in physical relations with physical events.

3.5 Incorrigibility and Imperceptibility

The incorrigibility and imperceptibility objections to the Identity Theory take the form of attempting to obstruct the Identity Hypothesis by producing predicates to show (by Leibniz's Principle) that $\phi \neq \phi$. Both these objections tend to be countered by suggestions that the intensionality of the predicates involved vitiate their application against the Identity Hypothesis. In other words, the Identity Theorist supposes that these objections do not worry him and that the objector can maintain his predications without troubling the Identity Hypothesis.

I will suggest that this is not the case: that the objections can be put in a form in which they directly oppose the Identity Hypothesis. And that since the plausibility of the Identity Theory is in inverse proportion to the plausibility of the objector's view, the Identity Theory cannot rest easy until it has actively discredited the objections. I will first discuss incorrigibility.

3.51 According to Baier, a mental event or state is 'something about which the person whose private state it is has final epistemological authority, for it does not make sense to say "I have a pain unless I am mistaken".' (Baier, p. 98).

Smart summarises this objection as follows: 'No physiological evidence, say from a gadget attached to my skull, could make me withdraw the statement that I have a pain when as a matter of fact I feel pain.' (Smart (3), p. 107). And Smart's reply to this is essentially that should a conflict arise (between introspective and external evidence), then he would have to reject the Identity Hypothesis. But as part of his theory Smart does not think that such a situation would in fact arise: 'It should be recalled that I put forward the brain-process thesis as a factual identification, not as a logically necessary one.' (Smart (3), p. 107).

But this alone does not parry Baier's objection. For Smart thereby admits the efficacy of a schema for the falsification of the Identity Hypothesis which relies on the incorrigibility of introspective evidence. Baier's objection should have been summarised by Smart as something like: 'It is logically impossible that physiological evidence could make me withdraw the statement that I have a pain when I feel pain'. And ϕ reports would not be subject to this."

Bradley has a different sort of counter to the incorrigibility objection. He forms an incorrigibility predicate (cf. Bradley, p. 174) for which I will use 'Ix' read as:

'It is logically impossible that x is judged by a person to occur as part of his experience and x does not occur'.

(The scope of 'It is logically impossible that' extends down to 'not occur'.) Bradley takes 'judges that' as 'explicitly formulates and assents to the proposition that'.[12]

Bradley now disputes that $(I\phi.\sim I\psi)\supset(\phi\neq\psi)$, on the grounds that the modal operator precludes the normal application of the predicate in Leibniz's Principle. Bradley takes what he regards as an analogous predicate, for which I will use 'J', such that 'Jx' is read as:

'It is logically impossible that x is not observable before dawn in the eastern sky'.

Then if 'y' and 'z' are replaced by 'the morning star' and 'the evening star' respectively (where these are construed not as names but as descriptions, e.g. 'the last heavenly body observable before dawn in the eastern sky' elucidating 'the morning star'), then $y = z.\sim(Jy\supset Jz)$. And so Bradley argues that analogously because of the intensionality arising from the modal operator in the predicate I, $\sim((I\phi.\sim I\psi)\supset(\phi\neq\psi))$. (Cf. Bradley, p. 175.)[13]

However, it does not <u>follow</u> from this analogy that the predicate I is intensional. The only way to <u>show</u> that a predicate F is intensional is to show that the truth of 'Fx' may depend on how x is described - that possibly $x = y.Fx.\sim Fy$. Clearly with a predicate such as J it is a simple matter to show that J is not a genuine property of the morning star, independent of how the morning star is described.

Certainly, Bradley casts suspicion on the use of a modal predicate such as I in connection with Leibniz's Principle. But it seems clear that the incorrigibility objector would insist that under <u>any</u> description of ϕ, $I\phi$: that, in other words, $(x = \phi)\supset Ix$. For he would be expected to take I as one of the <u>defining</u> properties of ϕ (in view of the fact that he takes (o) to require (p)). He just could not conceive how something could <u>be</u> one

of his mental events unless he had incorrigible introspective evidence of its occurrence.

I will return to this in a moment. For at this stage it might be supposed that if intensionality in I due to the modal operator is not proven, surely 'judged' in I is problematic.[14] For 'x is judged to occur' is normally an intensional context. So since $\sim((x = y) \supset (x \text{ is judged to occur} \equiv y \text{ is judged to occur}))$, the Identity Theorist need not be troubled by a failure in the Leibnizian substitution.

But we can get round this by reconstructing the incorrigibility predicate. And, while we are about it, we can acknowledge the qualification that the judgement should be sincere and free from linguistic error: for if 'judges that' is taken, with Bradley, as 'explicitly formulates and assents to the proposition that', $\sim I\psi$ would be possible, since the subject might not <u>sincerely</u> assent, or he might not formulate the proposition he means to. To deal with this we can modify the predicate to contain 'Asserts' where 'A Asserts' is read as:

> 'A sincerely, and with linguistic competence, explicitly formulates and assents to'.

Now, we can abandon any pretence that the occurrence of 'x' in 'A Asserts the proposition that x occurs as part of his experience' is purely referential, and form the family of predicates, I'_{ψ} for each ψ, such that 'I'_{ψ} x' is read as:

> '$\sim\Diamond$(A Asserts the proposition expressed by 'ψ occurs as part of A's experience' and x does not occur)'.[15]

In contrast to Bradley's predicate, we now only have one (a purely referential) occurence of the variable - on the assumption that the <u>modal</u> context doesn't block referentiality. And only the modal operator could be accused of rendering the fact that $\sim I'_{\psi}\,\phi$ inapplicable against the Identity Hypothesis. But against this it would be claimed by the incorrigibility objector that $(x = \psi) \equiv I'_{\psi}\,x$. And as I mentioned just now, it is no simple matter to

disprove this.[16]

Moreover the use of the predicate I'_{ψ} conforms to the intuitive idea behind the incorrigibility objection, that it is a genuine property of $\not\phi$ that its occurrence is not logically independent of the beliefs (etc.) of the subject. Whereas the occurrence of ϕ is logically quite independent of these beliefs.

So, the incorrigibility objector will argue:

(1a) $(x = \not\phi) \equiv I'_{\psi}\, x$;

(2a) $\sim I'_{\psi}\, \phi$;

so (3a) $\phi \neq \not\phi$.

But the Identity Theorist can insist that there <u>is</u> a description of $\not\phi$ such that I'_{ψ} is not truly attributable to $\not\phi$: a neural description. Thus he could insist:

(1b) $\not\phi = \phi$;

(2b) $\sim I'_{\psi}\, \phi$;

so (3b) $\sim((x = \not\phi) \equiv I'_{\psi}\, x)$.

There is genuine confrontation here between the Identity Theorist and his objector, as a result of which the Identity Theorist cannot allow (1a). But for the Identity Theorist to <u>dispose of</u> the objection, rather than merely survive by simply denying (rather than arguing against) (1a), he has to reject (1a) from some premiss or argument other than (1b). That's to say, in order to advance against the genuine problem facing his position - a problem in that there <u>are</u> intuitive justifications for (1a) - the Identity Theorist cannot rely on the fact that his objector cannot <u>prove</u> (1a). He must actively descredit (1a).

We should note also that unless the view that (o) requires (p) is explicitly disputed by the Identity Theorist, then in denying (1a) he will be vulnerable to the accusation that he is denying (o) - and that therefore he must be maintaining something like a II-Identity Theory. (Of course, this won't matter if he is maintaining a II-Theory in the first place.)

3.52 Baier has another characterising modal predicate for mental events: ψ is necessarily imperceptible by the senses (cf. Baier, p. 98).

It might be thought that problems of modality might be side-stepped by considering a simple non-modal predicate as in 'x is imperceptible to the senses'. But if 'perceptible' is taken as 'perceptible directly to the senses', it seems unlikely that any physical candidate for the Identity Hypothesis would be perceptible. While if 'perceptible' is taken as 'perceptible that', then not only does intensionality appear, but also it may be claimed that it may be perceptible <u>that</u> ψ occurs, in roughly the sense in which it is deducible that ψ occurs: there may be a theory contained in a view that ψ occurs iff ϕ occurs, by which it may be perceived or deduced <u>that</u> ψ occurs.[17]

Thus for an imperceptibility objection we do need the modality (and the extensional sense of 'perceptible'). Thus suppose 'Pyx' is read as 'y is a sensory means for directly perceiving x', and 'IMPx' is read as '$\sim\Diamond(\exists y)(Pyx)$'. The imperceptibility objector would then maintain that $(x = \psi) \supset IMPx$, but $\sim IMP\phi$, so $\phi \neq \psi$.

And again we have the position reached in the case of the incorrigibility objection: the Identity Theorist may bluntly argue that under the neural description 'ϕ' of ψ, $\sim IMP\phi$, and thus deny that $(x = \psi) \supset IMPx$. But this is not a constructive maneouvre, for clearly the modality of the imperceptibility predicate is just a symptom of the fact that the objector cannot conceive that a mental event could be a physical event - an event which could be perceived directly or which could causally engage with or otherwise be physically related to the events we do perceive directly. Really the Identity Theory must attempt to engage with the objector's justifications for this contention, rather than rely on the fact that the objector may not be able to prove his justifications are sound. For the Identity Theorist to counter through technicalities is rather like the case where a man is fighting a Hydra by neatly severing each head which attacks him.

Moreover, as in the incorrigibility case, to simply deny the imperceptibility premiss does not allow (o) without (p). An Identity Theorist would have to explicitly dispute the view that (o) requires (p) if he is to avoid being forced to hold a II-Theory.

3.6 The Location Objection

The location objections to the Identity Theory again take the form of attempting to produce an F such that $F\not\phi.\sim F\phi$.[18]

The location objections suggest that unlike ϕ, which has a spatial location (at least somewhere in the central nervous system), $\not\phi$ has no spatial location at all (though it is taken to have temporal location).[19]

3.61 (i) There is not in any sensation '<u>any element of actual locality, of inherent spatial order</u>, any tone as it were, which cries to us immediately and without further ado, "I am <u>here</u>", or "I am <u>there</u>"' (William James, pp. 157 - 158; he calls this denial of locality the "Local-Sign Theory"). (I will expand this objection below.)

(ii) Certain somatic sensations may tempt at least their "association" with physical locations; but there are mental events which do not suggest in any obvious way their, or any, location. A sudden thought, for example: 'Suddenly realising, in a chess game, that moving this pawn would endanger one's queen' (Malcolm, p. 171). Malcolm suggests that:

> ... in our ordinary use of the terms 'thought' and 'thinking', we attach no meaning to the notion of determining the bodily location of a thought. We do not seriously debate whether someone's sudden thought occurred in his heart, or his throat, or his brain. Indeed, we should not know what the question meant. We should have no idea what to look for to settle this 'question'. ((Ibid., p. 174.))

Bradley considers objection (i), suggesting a counter to the view that this objection shows a peculiarity of sensations not shared by the physical:

> what is it in the case of a physical <u>locandum</u> that cries, without further ado, 'I am here' or 'I am there'? ... So far as I know, nobody supposes that the marks by which we assign physical bodies to their respective places are specifiable without the use of a

> relational spatial vocabulary. Yet ... it is the fact that the marks by which we assign sensations to their places are not specifiable without the use of a relational spatial vocabulary that is invoked in support of the view that they are not located in the sense in which physical bodies are located. ((Ibid., p. 184.))

But it might be argued that what is at issue is not whether ϕ's can be ordered as can physical objects, by means of a relational spatial vocabulary; but whether the ordering results in those spatial relational properties which obtain between physical objects. In other words, although qualities of sensations may make them susceptible to a "spatial" ordering, is that ordering "in the same space" which contains physical objects? It may be appropriate for instance to speak of a weight as "located" in an ordering of weights, depending on their magnitude. Or of people "located" at a certain point in a personality spectrum. But this would not be locating in the space of physical objects. And a point that Lotze and James make is that the ordering of sensations (in a given sense modality), associated by 'the law of habit in the nervous system' (James, p. 158) with an ordering of sensed physical objects, is distinct from that ordering of physical objects (in physical space) which is achieved by the habitual correlation. This is a familiar point that can be illustrated by the analogy between a code of signs which have a certain (syntactic) ordering, and the items designated by the code, which may have a quite distinct ordering.[20] The syntactical properties of a certain expression in the code do not intrinsically point to the item "located" or designated by the expression - in the code this designation is arrived at by semantic association. And in the case of a certain sensation, again, it does not follow that it must locate a certain place or spatial relationship. For instance, on the basis of a certain visual image I may judge a banana to be to the right of an orange. But if my vision were to laterally invert, my judgement of location on the basis of that same visual image would alter. Similarly, if the kinaesthetic and tactile sensations associated with my little finger were exchanged for those associated with my third finger, I would have to alter the locations made

by the original set of sensations.[21]

But Bradley could reply to this rendering of the Local-Sign Theory as follows. There may be nothing about a somatic sensation which intrinsically suggests the actual location of the physical event or object that is sensed. But the Local-Sign Theory in that case does not obviously engage with the issue of locating the sensation (rather than some object or part of the body which is sensed). Bradley's point could still apply to this issue. Suppose, for instance, that I am engrossed in a book while travelling in an aeroplane. There may be nothing about the mental events I experience which enables me to tell that I am moving. There is nothing about them which suggests their location. But the same could be said of the interior of the aeroplane: there may be no clue as to the aeroplane's location which is provided by an examination of its interior. In the case of both the mental events and the aeroplane, I need a separate physical reference point to establish location.

This counter also applies to objection (ii). Once I have an external reference point, I can locate my head, and so I can locate a sudden thought which I may have.

But objection (ii) can resist this counter. For although we need another physical body to act as a reference point when locating a certain physical object, it's not clear that the process of locating a thought is exactly parallel. We do have the locutions involved in the phrases 'what is going on in Bill's head?', 'my thoughts are in my head not in my heart'. But do these locutions suppress, for the sake of convenience in non-philosophical activity, a location by convention? The following considerations tend to suggest that the mechanism by which, say, a sudden thought is located inside one's head, is rather different from that by which one physical object is located with respect to another:

(a) There is no "spatiality" about a sudden thought (or indeed a somatic sensation as distinct from the physical objects sensed). It does

not appear extended in space; it does not appear to occur in any volume of specific dimensions. Thus, though Bradley's counter applies to the point that an examination of a sudden thought (alone) does not reveal its location, it does not specifically counter this view: that there is no quality of a sudden thought by which it could be related to a physical reference point for location. In other words, though Bradley is right in suggesting that we should not expect to be able to locate a mental event without first establishing a physical reference point, it is arguable now that it would be of no help establishing a physical reference point since there is lacking the means to draw spatial (rather than conventional) relations between a mental event and that reference point. To support this view we can argue as follows:

(b) Suppose my skull contained my heart, and my brain was situated where the heart is normally - in my chest. Then there would be no indication of this from an examination of my mental events. Or suppose that the top of my skull is removed, so that while I am reading a book, say, my brain is being hoisted up and down (with appropriate extensions to all my nerve fibres). Again there would be nothing about my mental events which would enable me to locate them differently (there certainly shouldn't be - there would be no means of sensing anything unusual).

So there is no criterion by which we can begin to locate, say, a sudden thought, with respect to a physical locandum; though there are recommendations for the locution 'my thoughts are in my head, not my heart': the brain has been discovered to have a lot more to do with thoughts than the heart has, and the brain is inside the head. The brain is as a matter of fact inside the head, but my brain (or central nervous system) might be "radio-linked" to my body, while being stored separately from my body, for all I could tell from my mental events.[22]

3.62 The response to this sort of objection generally involves supposing that even if we can't determine the location of ϕ' through introspective evidence, we may nevertheless hypothesise that in fact ϕ' is where ϕ is: we may adopt a convention which assigns to ϕ' the location of ϕ.

This sort of defence involving location by convention is made by Shaffer, who notes Smart's anticipatory remark:

> we may easily adopt a convention (which is not a change in our present rules for the use of experience words but an addition to them) whereby it would make sense to talk of an experience in terms appropriate to physical processes. ((Smart (1), p. 62.))

Indeed, Smart's original reply to the location objection is ineffective, and would seem to require buttressing by something like location by convention:

> The after-image is not in physical space. The brain process is. So the after-image is not a brain process ... This is an ignoratio elenchi. I am not arguing that the after-image is a brain process, but that the experience of having an after-image is a brain process. It is the experience which is reported in the introspective report. Similarly, if it is objected that the after-image is yellowy-orange, my reply is that it is the experience of seeing yellowy-orange that is being described, and this experience is not a yellowy-orange something. So to say that a brain process cannot be yellowy-orange is not to say that a brain process cannot in fact be the experience of having a yellowy-orange after-image. ((Smart (1), p. 61.))

Smart's device here might work against the suggestion that ϕ is not yellowy-orange - since the experience is not yellowy-orange either;[23] but then the experience is no more evidently located than the after-image, as Bradley points out (ibid.,p.187).

The propriety of locating by convention is supported by Shaffer on the grounds that there is nothing about mental events that rules out their location by convention (Shaffer, p. 118). He says that there is nothing self-contradictory about speaking of mental events as located, whereas it would be self-contradictory to regard a fictional character as located, say, where the author created the character: fictional characters can be specified so as to rule out their being located. Location by convention in this case would be indefensible (except as an ellipsis). But he suggests that ψ cannot be specified so as to rule out its being located.

Cornman makes the bolder suggestion that we cannot argue that $\psi \neq \phi$ on the basis that only ϕ is located, because a predication to the effect that ψ has not that location would not have a truth-value: in effect it would be senseless rather than false to say that ψ has that location (Cornman (1), p. 129). Such predications are typical, Cornman suggests,

where tney involve a category mistake; and because Leibniz's Principle (requiring that the predications have a truth-value), would not apply in the case where the two terms are in different categories, 'it is at least possible that ... they both refer to the same thing' (ibid., p. 129).

But this position has the odd result that where on the face of it two descriptions are as different as they could be, they could nevertheless describe one and the same thing. It would then be possible for a certain fictional character to be a certain chair, since it would be senseless to make the relevant cross-predications. But from the fact that there is no criterion for determining the truth-value of 'Fx', it does not seem to me that it has to follow that it is senseless or truth-valueless to say that it is <u>not true</u> that Fx, or <u>not true</u> that x = y (where Fy). But even if it <u>were</u> regarded as senseless to say that it is not true that Fx, it would seem to me to follow that it should be <u>senseless</u> to say that x = y (again, if Fy), rather than <u>possibly true</u>.

Quinton argues that a convention for locating mental events <u>must</u> be adopted. He says that there are two possible answers to the question about the physical location of a mental event or state:-

> ... the dualist answer of Descartes and Hume (viz., nowhere) or the equally natural answer that they are where the people who are their owners or bearers are, that is, where their owners' bodies are or where the behaviour that manifests them or their proximate physical causes are to be found. In favour of the former choice is the fact that introspection does not appear to yield any spatial information whatever about its objects. ((Quinton, p. 335.))

Quinton then poses a difficulty for the former choice - a difficulty in individuating mental events or states:

> Suppose that two people, A and B, undergo qualitatively indistinguishable sentations that are strictly temporally coincident, beginning and leaving off at the very same moment. How are we to justify the belief, which we should certainly hold in these circumstances, that there are two sensations going on here and not just one? It is part of the hypothesis that the sensations of A and B are not distinguishable from one another in respect of their temporal and other introspectible characteristics. ((Ibid., p. 335.))

He concludes that if the mental events in question are not regarded as physically located, there would be no means of individuating them - for they

cannot be individuated either by their distinct possible physical causes in the two bodies (since one event can have two effects).[24]

Quinton strengthens his example by considering the supposition, conceivable to the dualist, that A and B are disembodied (eliminating publicly observable distinguishing features). He concludes:

> These arguments are intended to show that we must at least admit that mental events are spatial in the comparatively weak sense of being where the bodies of the people to whom they occur and by whom they are manifested are. ((Ibid., pp. 336 - 337.))

But Quinton's argument begs the question. For although one might sympathise with the conjecture that A's and B's sensations are indistinguishable, A's and B's sensations, in the dualist sense of 'sensation' do not admit direct comparison. For they are in principle private to A and B; so there would in principle be no criterion for comparing them. We can only compare A's and B's public parts - their bodies and behaviour. It would only be sensible to say that A's and B's bodies or behaviour are indistinguishable. So Quinton can only set up his example by assuming that the mental events in question are not the sort of events whose unlocateability he is trying to refute. Of course, there are <u>other</u> familiar arguments to attack the view of mental events which the dualist requires - and Quinton's argument might be a symptom of a sympathy with these other arguments for the discrediting of such entities. But unsupported by something like the private language discussion, Quinton's counterexample fails to counter the required view.[25]

<u>3.63</u> Malcolm criticises Shaffer's proposal for locating mental events by convention on the grounds that the stipulation that a thought is located in a certain region of the body is merely an elliptical way of speaking of the corresponding bodily processes (cf. Malcolm, pp. 174 - 175, fn.).

Malcolm argues further:

> Suppose we had determined, by means of some instrument, that a certain process occurred inside my skull at the exact moment I had the sudden thought about the milk bottles. How do we make the further test of whether my <u>thought</u> occurred inside my skull? For it would have to be

> a <u>further</u> test: it would have to be logically independent of the test for the presence of the brain process, because Smart's thesis is that the identity is <u>contingent</u>. But no one has any notion of what it would mean to test for the occurrence of the thought inside my skull <u>independently</u> of testing for a brain process. ((Ibid., p. 175.))

I cannot think of any uncontroversial and self-evident principle which would require that we would have to assign locations to mental events independently of neural events for the Identity Hypotheses to be contingent.[26] The trouble with Malcolm's remarks as an objection to Shaffer is that they amount to little more than a restatement that the location of mental events at an appropriate place would have to be by convention. However much sympathy we might feel for Malcolm's misgivings, they do not destroy Shaffer's hope that, although we have no means of ascertaining the location of a mental event except as determined by the "corresponding" bodily processes, we can be satisfied by an assignment of location which is dependent on these latter processes.

Shaffer hopes that the location objection is seen as a difficulty in actually <u>determining</u> the locations of mental events: they can't be located independently of the claims of the Identity Theory. But since (he argues) it is not self-contradictory to suppose that, in spite of this difficulty, mental events <u>are</u> located, we can without logical difficulty suppose that they are appropriately located.

The issue clearly hinges on whether or not it is self-contradictory to suppose that mental events are located. While it is the case that X <u>might</u> be at a place s, it can be useful in certain contexts to locate X at s by some convention. Shakespeare might very well have been born in a certain house in Stratford; so the convention that a certain house in Stratford <u>is</u> the birth-place of Shakespeare is acceptable. But it would not be so acceptable to propose that $\not\mu$ is at s by convention, should $\not\mu$ not be the sort of "thing" which <u>could</u> be at any place.

And clearly the dualist objector will maintain that it would be self-contradictory to suppose that $\not\mu$ could be spatially located. For on the basis of the considerations I mentioned as intuitively justifying the

objection (ii) of 3.61, the following general objection would be proposed by the dualist:

Suppose S is some physical spatial relation, and 'S'x' is read as '~◇(∃y)(Syx)'. Then (the dualist would argue) $(x = \not\phi) \supset S'x$, but $\sim S'\not\phi$, so $\not\phi \neq \not\phi$.

And as with the incorrigibility and imperceptibility objections, it is not constructive for the Identity Theory merely to deny the first premiss - a denial which is implicit in the "location by convention" defence.

3.7 Explanation and the Identity Theory

I will try to illustrate in a more concrete manner the difference between the I- and II-Identity Theories. (To do this I will use an example very like one I used in 2.43 in connection with the Materialist Compatibility Point.) And we shall see that a general objection to the I-Identity Theory can be put as a demand for an explanation where none is given.

3.71 Suppose that there exists a community of organisms which exhibit behaviour which matches ours; indeed, they would be indistinguishable from ourselves as a group, both behaviourally and physically, if it wasn't for some minor physical difference. Now:

(i) A certain scientist sets about finding a physical explanation for their behaviour. Eventually he establishes blueprints of their physical structure, and forms a complete and successful theory which explains all their physical states and behaviour (in physicalist terms). Now, he believes that these organisms are sentient and conscious, etc. - quite naturally, for he converses with them, and they describe themselves as sentient and conscious; and he has no reason to think they are significantly different from himself as regards their "mental life". And he believes that there occur in each of these organisms mental events which they can report as occurring in themselves and of which they can be introspectively aware. But it seems to the scientist that these mental events are in principle imperceptible to him.

(He believes that he cannot perceive their pains (etc.) - that their pains are none of the physical events which he can (directly or indirectly) perceive as occurring while they are in pain.)

He is inclined therefore (a) to regard these mental events as occurring in addition to the physical events occurring in these organisms; (these latter events he can monitor and explain perfectly). However, a philosopher suggests to him that instead of maintaining this uneconomical ontology he could usefully (b) regard these two streams of mental and physical events as being as a matter of fact one and the same: - take this as a scientific hypothesis.

The scientist notes that if he kept to the alternative (a), he would have to accept the existence of events for which he had no scientific explanation (since he can only explain in scientific terms the (physical) events which he is able in principle to perceive). So he espouses the view suggested in (b), thinking that he would then have a scientific explanation for all events, in particular, that he would have a scientific explanation for the mental event ψ, since he has a scientific explanation for the physical event ϕ, and $\phi \neq \psi$.

(ii) Another scientist has it on the highest authority that these organisms are actually automata. He supposes that they aren't conscious at all, but are just very complicated machines - very complicated lumps of matter.

He also has a blueprint for each of these organisms, and the full physical theory explaining the physical states and behaviour of these organisms. So in this respect he has the same information as the first scientist: he has a physicalist explanation for all their behaviour (overt and covert), including of course lip movements, "verbal" behaviour, and so on.

These organisms have "grown up" in the company of human beings, and have learnt our natural language (say, English). So the second scientist (like the first) has a scientific explanation for each language learning

step, output, etc. in the case of these organisms. So he has a physical explanation for the occurrence of each syntactic output for these organisms, and can explain the occurrence (qua syntactic/physical entities) of terms like 'me', 'dog', 'pain', 'after-image', 'mind', in statements made by these organisms.

Now in the case where one of these organisms utters, in the normal course of things, 'my left foot is bleeding', the scientist has no difficulty in interpreting the meaning of the expressions involved: e.g., 'my left foot' refers to the organism's left foot - the expression is used by the organism just as we would use it. But although the scientist can explain every physical stage involved in the production of the output by an organism: 'I have a pain in my foot', he is not sure how to interpret the output. He decides eventually that the organism is asserting the occurrence of a certain physical event inside him. Of course, he isn't even tempted to take such an output as asserting the occurrence of a non-physical event inside the organism, for he has it on the highest authority that these organisms are lumps of matter and nothing more.

So, having found suitable internal physical processes with which to implement his view, the scientist proposes that an event ψ asserted to occur by the organism in a certain sort of utterance is a physical event ϕ occurring in the organism. So the truth-conditions of a sentence containing an expression 'ψ' purporting to refer to an internal event in the organism are adequately accounted for if 'ψ' is taken to refer to ϕ.

Now, the scientist in (ii) can be aligned with the II-Identity Theorist (if we put ourselves in place of the organisms in the example). Should this scientist find that his Identity Hypothesis '$\psi = \phi$' proves problematic, he could simply abandon the view that 'ψ' works as a genuine referring expression in the language moves containing it. If his Identity Theory is false, an ontological dualist theory of the organisms is not a consequence. It does not follow on his view that if his Identity Theory proves to be false, and

it is not the case that 'ψ' refers to ϕ, then there must be the two events ψ and ϕ occurring.

But the scientist in (i), who can be aligned with the I-Identity Theorist, is in quite a different position. For he accepts that both 'ψ' and 'ϕ' refer to events whether or not $\psi = \phi$: he allows that ψ may be additional to ϕ, but suggests that as a matter of fact it isn't. If his Identity Hypothesis fails, he then has the two events ϕ and ψ. Moreover, the scientist in (i) accepts that unless he can maintain that $\psi = \phi$, he cannot scientifically explain the nature and occurrence of ψ.

The scientist in (ii), on the other hand, is quite satisfied that he has already a scientific explanation for every event occurring in these organisms. It is not for the purpose of accounting for events that would otherwise go unexplained that the second scientist maintains that $\psi = \phi$. Putting it crudely, it is not mental events he has to explain but 'mental events'. That's to say, he wants an interpretation for terms like 'mental event' in language moves: - he already knows the "physics" of their use, as he might know why, electronically, a computer outputs certain symbols, e.g., 'P-K4'. The second scientist wants a semantic interpretation of some of the expressions involved in the language moves of these organisms. (In the case of the computer output, we could give an interpretation, if the context were right, such that the output described, or made, a move in a chess game.)

Now, the first scientist initially has a problem: ψ seems to him to be additional to physical events. Indeed, he accepts that mental events may (logically possibly) be additional to physical events. But of course the hypothesis that $\psi = \phi$ cannot explain away his problem: it cannot explain why what seems to him to be and may be additional to physical events is not additional. The mere supposition that mental events are not additional to physical events can't explain why they seem to be but aren't.

It might be supposed that there may be a scientific factual explanation

of this problem. But all the scientist's physicalist theory could do was explain the occurrence of all the physical events: indeed, it was supposed that the scientific theory was compatible with both the monist and the dualist situation.

The only possible sort of explanation available for the scientist's problem would be one which showed that his reasons for taking mental events to be additional to physical events were wrong or confused. But the scientist would not look for this sort of explanation while he maintained that it was logically possible that mental events were additional to physical events.

3.72 There are two points we can make here. In the first place, an explanation is needed for the I-Identity Theory, and not given by it, to explain why what are prima facie additional to physical events are not additional. And secondly, the I-Identity Theory is shown to be unstable in connection with this demand for an explanation. For if mental events seem to be additional to physical events, and logically possibly may be, then an explanation why they seem to be additional but aren't can only be a factual scientific explanation (of the sort which explain why lightning is an electric discharge, or the boiling of a liquid is its vapour pressure equalling atmospheric pressure). But a scientific factual explanation is ruled out also, since scientific theory is taken to be compatible with both dualism and monism.

This sort of impasse is evident in Smart (in his I-Theory mood), where the main justification for the Identity Theory against dualism is the parsimonious nature of the former, and no further explanation is given or looked for.

And this brings us to another general instability present in the I-Identity Theory: it allows coherence to the rival dualist, while lacking the empirical bias on the basis of which scientific theories are preferred to their rivals. But then the Identity Theorist is going to be eternally at loggerheads with the dualist - for in allowing coherence to dualism he is

allowing coherence to arguments which render the Identity Theory, though more economical, quite implausible.

Of course, Smart's views cannot be presented as suggesting a pure I-Theory, or a pure II-Theory. Moreover, Smart in places does argue as a more satisfactory Identity Theory should: against certain underlying features of the dualist's view. But the inherent weakness and ambivalence of the original conception of the Identity Theory tend to blight the discussions of it.

It is not enough, of course, to show dualism as incoherent - the PLA does that, to my mind. What has to be done, as anticipated, is settle the question of whether we have both to abandon (p) and (o), or whether we can manage to retain (o) without (p). The II-Theory would be under some obligation to undermine the conviction that (o) obtains. But I think a nicer view to aim for is one in which we can allow introspective awareness, and so on, while maintaining that we are purely physical - a monistic account which cannot be accused of "blank automatism".[27]

I will now briefly discuss one argument against and one argument for the Identity Theory, neither of which superficially fits into the standard strategies outlined. But we will find that the active ingredients of these arguments are familiar, and allow the arguments to be usefully placed in the context I have outlined.

3.8 Kripke's Argument Against the Identity Theory

This argument appears to be based on Kripke's view of identity statements. Its form can be rendered by the following sketch:

(a) 'Hesperus is Phosphorus', 'Gold has atomic number 79', 'heat is molecular motion' are all necessarily true, if true. (The view that proper name identity statements are necessarily true, if true, is perhaps less controversial than the view that certain identifications and predications involving terms for natural kinds ("common names") should be construed likewise.)

(b) As a result the following statements are, at the very least, infelicitous:

'Hesperus might not have been Hesperus';

'Heat might not have been heat';

'Hesperus might not have been Phosphorus';

'Heat might not have been molecular motion'.

(c) However, we have available in these instances a way of saying what we perhaps wanted to say by the above statements, which does not contain inaccuracy of expression. The states of affairs whose possibility motivated the above statements can be described by recasting the statements without misuse of the designators.

(d) It is essential that the "illusion" of contingency regarding, say, 'heat is molecular motion', be satisfied by the device of recasting the statement so as to express the fact that certain other states of affairs are possible.

(e) There is an obstructive disanalogy between 'heat is molecular motion', say, and 'pain is the stimulation of C-fibres' (Kripke's favoured version of a typical hypothesis in the mind-brain identity theory). The illusion of contingency, which is based on a proper need to describe certain possible states of affairs, cannot be satisfied in the case of 'pain is the stimulation of C-fibres'.

I don't here want to go into Kripke's views on identity statements. But it seems to me that his suggestions regarding identity statements are neutral with respect to the mind-brain issue, and that we can discern in his argument a familiar dualist bias. The Identity Theorist can thus direct a counter at this inherent dualist bias, without having to settle the status of the relevant identity statements.

3.81 Kripke's device for explaining the illusion of contingency goes as follows. He is talking about a table in front of him, which is made of wood:

> What, then, does the intuition that the table might have turned out to have been made of ice or of anything else, ... amount to? I think

> it means simply that there might have been a table looking and feeling just like this one and placed in this very position in the room, which was in fact made of ice. In other words, I (or some conscious being) could have been qualitatively in the same epistemic situation that in fact obtains, I could have the same sensory evidence that I in fact have, about a table which was made of ice. ((Kripke, p. 332.))

So Kripke suggests:

> When someone says, inaccurately, that heat might have turned out not to be molecular motion, what is true in what he says is that someone could have sensed a phenomenon in the same way we sense heat, that is, feels it by means of its production of the sensation we call 'the sensation of heat', ... even though that phenomenon was not molecular motion. ((Ibid., p. 338.))

But Kripke argues that it is not possible to explain away analogously the feeling that the identification of pain with the stimulation of C-fibres, if it turned out to be a factual discovery, could have turned out otherwise:

> In the case of the apparent possibility that molecular motion might have existed in the absence of heat, what seemed really possible is that molecular motion should have existed without being felt as heat, that is, it might have existed without producing the sensation of S, the sensation of heat. ((Ibid., p. 339.))

But

> is it analogously possible that a stimulation of C-fibers should have existed without being felt as pain? If this is possible, then the stimulation of C-fibers can itself exist without pain, since for it to exist without being felt as pain is for it to exist without there being any pain. Such a situation would be in flat out contradiction with the supposed necessary identity of pain and the corresponding physical state. (Ibid., p. 339.))

So, according to Kripke, pain could not be the stimulation of C-fibers, since it would always be accurately said that pain might not have been the stimulation of C-fibers.

The disanalogy which Kripke tries to pick out is that, in the case of heat, a qualitatively identical epistemic situation to one in which an observer "apprehends" heat, need not have been one in which the observer is in fact apprehending heat. Whereas an epistemic situation qualitatively identical to one in which an observer has a sensation S simply is, he argues, one in which an observer had that sensation. Kripke has the idea that, in the case of heat and molecular motion, there is something, the sensation of heat, which is "intermediary" between the external phenomenon

and the observer, by means of which a felicitous recasting can be constructed to express the alternative possibility:

> In the case of molecular motion and heat there is something, namely, the sensation of heat, which is an intermediary between the external phenomenon and the observer. In the mental-physical case no such intermediary is possible, since here the physical phenomenon is supposed to be identical with the internal phenomenon itself. ((Ibid., p. 339.))

3.82 Certainly, feeling pain could not be an intermediary between the observer and the stimulation of C-fibres. But I don't think Kripke can use this feature to object to the identification unless he espouses an explicitly dualist view.

If pain is the stimulation of C-fibres, we must not allow as possible, according to Kripke, that the stimulation of C-fibres should have occurred without there being pain. But we must be able to say that it is possible that the stimulation of C-fibres should have occurred without being felt as pain - analogously to the way in which we are able to say that it is possible that molecular motion should have occurred without being felt as heat. Kripke then argues that since to <u>feel pain</u> is for there to be pain, if it is possible that something should have occurred without being felt as pain, then it is possible that it should have occurred without there being pain.

But it may be argued that all this shows is that <u>in the sense of 'feel'</u> in which we can allow that to <u>feel pain</u> is for there to be pain, the statement 'it is possible that the stimulation of C-fibres should have occurred without being <u>felt as pain</u>' is <u>not analogous</u> to the statement 'it is possible that molecular motion should have occurred without being felt as heat'. In other words, that the sense in which molecular motion is felt as heat is a genuine epistemic sense in which an observer, through an intermediary (the sensation of heat), apprehends heat. Whereas the sense of 'feel' in which to <u>feel pain</u> is for there to be pain, is the sense in which we feel blue or feel hungry - and is not the epistemic sense in which <u>something</u> is felt as, taken as on the basis of sensory evidence, heat. And that in the

sense in which to feel pain is for there to be a pain, neither the stimulation of C-fibres nor the pain is epistemically felt: since the subject is not in a genuine epistemic situation in which he apprehends the pain. There is no genuine sense in which there is the pain and the observer who observes the pain, in contrast to the sense in which there is heat, or molecular motion, and the observer who feels heat (note: heat, rather than the sensation of heat).

But for Kripke to execute his argument against the C-fibre identification, he has to show not that recasting is unsuccessful using the sense of 'felt as pain' which is disanalogous to that of 'felt as heat', but that there is not available a sense of 'felt as pain' which is analogous to 'felt as heat', through the use of which we can successfully recast.

In other words, we wouldn't want to recast by means of the sense of 'feel' in which to feel pain is for there to be pain, for that is not a genuine epistemic sense: there is no observer apprehending an object, there is just a person-in-pain.

Nor do we need to recast by means of this sense of 'feel'. There is no intermediary between the observer and the stimulation of C-fibres/feeling of pain in the case where the "observer" is the person-in-pain, but then the person-in-pain is not an observer of the pain. What we need is a genuine observer, who can be in a genuine epistemic situation with respect to the feeling-of-pain - in which an observer can "feel" the feeling-of-pain in a sense of "feel" analogous to the sense in which he may feel molecular motion. Such an observer would be someone other than the person-in-pain. And if pain is the stimulation of C-fibres, we could recast by saying, for instance, that it is possible that an observer should have been in an epistemic situation qualitatively identical to one in which on the basis of sensory evidence he took a feeling-of-pain to occur (in some other person), but there was no stimulation of C-fibres (in that other person). Or, the other way round, it is possible that a stimulation of C-fibres should have occurred (in person A) without being "felt", taken on the basis of sensory evidence

by an observer (B), as a feeling-of-pain (in A).

The following is an unfair rendering of the flaw in Kripke's argument, but it is useful nevertheless.

Suppose that a (purely physical) robot R exhibited pain-behaviour. We may be reluctant to say that it felt pain or felt anything. But suppose we did nevertheless describe R as feeling R-pain. For R to feel R-pain is for there to be R-pain in R.

Now suppose it was maintained that R-pain was the activation of C-circuits. Then it must be possible to say that the activation of C-circuits could have occurred without being felt as R-pain. But if this is possible, then the activation of C-circuits could itself occur without R-pain, since for it to occur without being felt as R-pain is for it to occur without there being any R-pain.

Of course, Kripke would maintain that the robot didn't feel anything, whereas we do feel something when we feel pain.

But what this unfair rendering gets across is this: for the activation of C-circuits to occur in R is just for that physical process to occur in the machine. We don't have to suppose that there is a physical observer in the machine which has to ascertain that that process is occurring. In order to recast 'R-pain might not have been the activation of C-circuits' we need to put it in terms of a genuine observer who is observing R.

Similarly, it could be argued that the feeling of pain in a person, the hurting, only needs to occur to be a hurting: this event does not, as well as occurring, need to be apprehended by a mental subject who is in an epistemic relation with respect to the hurting.

And now this brings out the point Kripke would have to and no doubt would insist on: that when a person feels pain (in the colloquial sense), he really is in a genuine epistemic situation in which he apprehends the pain - an introspective epistemic situation, but an epistemic situation nonetheless. Moreover, Kripke would have to and no doubt would insist that

to be in an introspective epistemic situation qualitatively identical to one in which a person apprehends pain is for the pain to occur.

This now is recognisable as a standard dualist contention: one in which a mental subject is logically tied to his object of awareness. And this is the active ingredient in Kripke's argument against the Identity Theory. For then he could claim that 'introspectively felt as pain' is analogous to 'felt as heat' - and he could then maintain that if it is possible that the stimulation of C-fibres should have existed without being introspectively felt as pain, then it can itself exist without there being pain.[28]

3.9 Davidson's Argument For the Identity Theory

Davidson argues that despite the differences between mental and physicalist language, individual dated mental events must be describable in physicalist terms.

Davidson describes his position as anomalous monism:

> Anomalous monism resembles materialism in its claim that all events are physical, but rejects the thesis, usually considered essential to materialism, that mental phenomena can be given purely physical explanations. Anomalous monism shows an ontological bias only in that it allows the possibility that not all events are mental, while insisting that all events are physical. ((Davidson (1), p. 87 - 88.))

According to Davidson, mental events, so described, are not susceptible to law-like generalisations; and there is no important sense in which psychological discourse is reducible to physicalist terms. Nevertheless, individual dated mental events can, and on certain assumptions must be individual dated physical events.

Davidson argues as follows:

(i) Mental events, at least some mental events, interact causally with physical events;

(ii) where we say that x causes y, there must be a strict covering law which x and y instantiate;

(iii) there are no strict laws which apply to mental events, so

described; (Davidson describes this as the anomalism of the mental). Davidson discusses the concepts required to implement his argument; of these I will question below only his notion of a mental event, which he gives, more or less, as follows: first mental verbs are roughly characterised; these are psychological verbs as used when they create apparently non-extensional contexts. Then a mental event is characterised as one which requires the essential use of at least one mental verb in its description. As Davidson admits, his test for the mental discards privacy in favour of Brentano's intentionality. And Davidson also notes problems with his test as a suitable and satisfactory spotlight on mental events. But in a way this doesn't matter for Davidson, since the form of his argument is really to show that the sort of event terms that could replace 'x' and 'y' in 'x caused y' must refer to physical events. So Davidson in effect hopes to use '... caused...' like a vacuum-cleaner - any event term which found itself used in one of the two slots in '... caused...' would be sucked into a "dust-bag" containing just terms referring to physical events. So he hopes that should anyone present him with a term purporting to refer to a mental event, the "vacuum-cleaner" would eat it up.

So Davidson argues:

> Suppose m, a mental event, caused p, a physical event; then under some description m and p instantiate a strict law. This law can only be physical... But if m falls under a physical law, it has a physical description; which is to say it is a physical event. An analogous argument works when a physical event causes a mental event. So every mental event that is causally related to a physical event is a physical event. In order to establish anomalous monism in full generality it would be sufficient to show that every mental event is cause or effect of some physical event; I shall not attempt this. (((1), pp. 99 - 100.))

Thus Davidson exploits the non-extensionality of explanatory laws; which allows 'ψ' to refer to ϕ even though ψ under this description may not be susceptible to law-like generalisations about it. And admitting that ψ can cause and be caused by physical events implies, according to Davidson, that there is a description of ψ to which strict laws can apply. So since the only strict laws available for complementing the causal relations involved are physical, there

must be a description of ~~it~~ suitable for instantiating physical laws.

Davidson illustrates the anomalism of the mental with an example similar to the one I used in connection with the Materialist Compatibility Point (cf. 2.43). He imagines that l'homme machine has been synthesised to be physically and behaviourally indistinguishable from a man (Davidson (2)). And he notes the radical difference between what goes on in saying of this synthesised chap, Art, that he intends, desires, believes, etc., on the one hand, and giving a physicalistic account of the changes in his body on the other. The two accounts do not readily mix. But this is compatible with only physical events happening in Art.

Davidson notes that some degree of dependence between mental and physical events must obtain:

> Such supervenience might be taken to mean that there cannot be two events alike in all physical respects but differing in some mental respect, or that an object cannot alter in some mental respect without altering in some physical respect. (((1), p. 88.))

But he emphasises that dependence or supervenience of this kind does not entail reductibility through law or definition. And Davidson provides an analogy for this sort of dependence without reducibility: truth in a formal language cannot be reduced to syntactical properties, and yet each true sentence in the language can be referred to by its syntactical properties.

3.91 An illustration of Davidson's general position, and one which retains the interpretation/syntax contrast present in his analogy, can be achieved through an example involving a machine playing chess.

Let's suppose that I'm playing chess with a suitably programmed computer. We can regard chess moves in the example as analogous to mental events - I will in places call chess moves chess events. And we can regard the hardware events (say, electronic) involved in the machine's internal processes and output production as analogous to the general class of physical events in Davidson's view.

Now, I would be strongly inclined to admit causal interactions between

chess events, and between chess and physical events. For instance, I might say any of the following (my moves are taken to be inputs in the machine, for simplicity):

'My second knight move caused my opponent to fianchetto his bishop.'

'My repeated check caused my opponent's output 'I offer a draw'.

'My rook sacrifice caused my opponent's time-overload light to come on'.

The position analogous to Davidson's for this setting is to suppose - very plausibly - that all my opponent's particular dated chess events are physical events. In support of such a view we should note that all the "outputtings" of the machine are physical events, have a physical description; but that these events are interpretable as constituting moves in a game of chess: a certain output can be described syntactically (literally), and also as a move in a chess game.

Moreover, I could give, in explanation of my opponent's fianchetto, some explanatory account involving weaknesses and strengths of chess positions. But there may be no strict law-like generalisations which his chess moves (so described) instantiate. (There may be some law-like statements about chess, by way of rules, for example.) On the other hand, there would be strict laws governing the hardware events in my opponent: my opponent's chess moves described as physical events can be explained in terms of laws of electronics, etc.. The occurrence of any hardware event (so described) in my opponent will instantiate a strict hardware law.

Now, analogously with Davidson, we could argue that there is a physical description of the individual dated move of my opponent's: 'Bishop takes queen' in our second game. But there may be no predicate of electronics "corresponding to" '... is a queen capture'. Compare this with Davidson's:

> What I have supposed is that for any particular, dated psychological event we can give a description in purely physical terms; and so for any given, finite class of events, we can set up a correlation between psychological and physical descriptions. But although this can be done, it does not follow that such psychological predicates as 'x desires his neighbour's wife', or 'x wants a kaffee mit schlag',... which determine, if not infinite classes, at least potentially infinite ones: - it does not imply that such predicates have any nomologically corresponding physical predicates. ((2), p. 712 - 713.))

Moreover, even if a sceintist were able to physically describe, explain and predict every physical event occurring in the chess machine, it does not follow that he could describe, explain or predict a single chess event (so described) occurring in the chess machine: for the simple reason that the scientist may know nothing about the game of chess - about that background against which to describe the physical events in the machine.[29]

Compare Davidson's:

> Even if someone knew the entire physical history of the world, and every mental event were physical, it would not follow that he could predict or explain a single mental event (so described, of course). (((1), p. 100.))

Davidson suggests:

> There is no important sense in which psychology can be reduced to the physical sciences. (((2), p. 722.))

Similarly, it could be claimed that chess is not reducible to electronics (or to the physical sciences generally). (Even if a winning algorithm for chess were discovered, this would just be reducing how to win in chess to a recursive procedure.)

But all this is clearly compatible with an ontological monism regarding the events which occur in a chess machine.[30] And in as much as causal relations might be said to obtain between chess events, the laws which moves would then instantiate would have to be physical laws: since there would be no strict laws of chess by which one could predict or generalise from responses in an actual game (which might of course contain "bad" moves, unexpected moves, etc.). So if we were to say that a certain chess event caused or was caused by some other event, this would require us to admit, according to Davidson, that under some description that chess event is susceptible to law-like generalisations of an appropriate sort. And it doesn't matter if we don't know the laws operating on chess moves qua physical events (as would be the case when playing against a chess machine with an unknown program) - Davidson's argument suggests that we would imply, when making the causal statements, that there were discoverable strict laws

explaining the sequence of events:

> If one event causes another, there is a strict law which those events instantiate when properly described. But it is possible (and typical) to know of the singular causal relation without knowing the law or the relevant descriptions... it is possible to know that a mental event is identical with some physical event without knowing which one (in the sense of being able to give it a unique physical description that brings it under a relevant law). (((1), p. 100.))

3.92 Now, Davidson's position appears to me to be strongly aligned with the II-Identity Theorist's view.

For Davidson's criterion of the mental is geared to recognising a certain sort of expression (rather than a certain sort of event). In this respect it is a development of Smart's suggestion, in his II-Theory mood, that what is important in characterising sensations is being able to recognise the class of sensation reports (cf. 3.3). Then, having picked out "mental" terms, Davidson provides a persuasive argument for the view that 'ψ' must refer to some ϕ (where in this case 'ψ' and 'ϕ' are individual event terms): the truth conditions for statements of the form 'ψ caused X', 'X caused ψ', can only be satisfactorily accounted for by having 'ψ' refer to some ϕ, where X and ϕ instantiate an appropriate causal law.

But as far as I can see, Davidson only illustrates how ontological monism is compatible with conceptual dualism - essentially by making a form of the Materialist Compatibility Point. His arguments as they stand do not refute ontological dualism (though he could back up his arguments by appealing to something like the Wittgensteinian line) - indeed, his arguments cannot really be taken to engage with the problem of events which seem to be private: the events which seem to resist an ontologically monistic account. For the force of Davidson's argument involving causal statements can be comparatively easily diverted by an ontological dualist who supposes that there are private mental events.

The dualist could suggest that when we say that ψ caused X (or vice versa), where X is a physical event and ψ is a private mental event, this involves an ellipsis. These causal statements should be interpreted something

like this: when we say that μ caused X, this is to say that we suppose that there is some physical event φ correlated with or "parallel" to μ which caused X. So in these contexts 'μ' is doing the work of 'some physical event correlated with the private mental event μ'.[31]

A not exactly analogous but related example might be the case where one says: 'the input π caused the machine to output zero', where in this case 'π' is elliptical for and doing the work of 'the numeral, or the physical (electronic) designation for π' - since of course the number didn't engage causally.[32]

Davidson admits that 'if there are mental events that have no physical events as causes or effects, the argument will not touch them' ((1), p. 80). But then it is only these events that could cause the ontological problem.

So, given the incompatibility of 'μ caused X' and 'the connotation of 'μ' is such that it could not cause X', the dualist can and has to abandon the view that in 'μ caused X', 'μ' is referring to the private mental event μ. To cut off this line of escape Davidson would have to address himself to this resistance to ontological monism: the resistance involved in supposing that the mental events we are introspectively aware of are (in principle) private.

Davidson does not of course tackle the problem of mental events of which we may be introspectively aware - at least not specifically - since he relies on the fact that these will be among the events described through mental discourse. Davidson is able to avoid discussing the supposed privacy of mental events - the privacy of introspected mental events - since he does not need the notion of privacy in order to recognise descriptions of mental events. And he perhaps regards the concept of privacy as just one of those concepts, found in mental discourse, which form a mentalistic scheme conceptually distinct from (and irreducible to) physicalist discourse - but which do not trouble ontological monism. But the supposed privacy of mental events is a concept in the mentalistic scheme which threatens to obstruct Davidson's

otherwise successful marriage of conceptual dualism to ontological monism: - since private mental events would not be picked up by Davidson's "vacuum-cleaner": '... caused ...'.[33]

3.10 The Identity Theory - So Far

So far the indications are that though the Identity Theory has not been disproved by the dualist, it doesn't look as though it has solved any ontological problems of mind.

The pressure to maintain (o) and (p) weighs against both the I-Identity Theory and the II-Identity Theory, and will weigh against any view which is incompatible with ontological dualism but fails to undermine either the thesis that (o) obtains, or the thesis that (o) requires (p).

The II-Identity Theory seems to be a fairly stable, internally untroubled view. But the pressure for (o) makes it intuitively difficult to accept.

And the pressure for (o) and (p) make the I-Identity Theory difficult to accept. Moreover, quite apart from this difficulty, the I-Identity Theory has internal difficulties, essentially arising from the fact that it allows that a logically incompatible view (ontological dualism) is logically possible.

The sort of Identity Theory I wish to espouse, retaining (o) without (p), must undermine the dualist's resistance - must show that we can have (o) without (p). This resistance is manifested in the objections to the Identity Theory which we have been considering, and which I will now summarise.

(i) Using the incorrigibility predicate (cf. 3.51):

'Asserts' is read as 'sincerely, and with linguistic competence, explicitly formulates and assents to'.

'I'_{ψ} x' is read as '$\sim\Diamond$(A Asserts the proposition expressed by 'ψ occurs as part of A's experience' and x does not occur)'.

There is a family of predicates, I'_{ψ} for each ψ, such that, it is argued,

$(x = \psi) \equiv I'_{\psi} x$,

$\sim I'_{\psi}\ \phi$,

so $\phi \neq \psi$.

(ii) Using the imperceptibility predicate (cf. 3.52):

'IMPx' is read as '$\sim\Diamond(\exists y)$(y is a sensory means of directly perceiving x)'.

Then it is argued that

$(x = \psi) \supset$ IMPx,

$\sim$IMPϕ,

so $\phi \neq \psi$.

(iii) Using the unlocateability predicate (cf. 3.63):

S is some physical spatial relation, and 'S'x' is read as '$\sim\Diamond(\exists y)$(Syx)'.

Then it is argued that

$(x = \psi) \supset$ S'x

$\sim$S'ϕ,

so $\phi \neq \psi$.

(iv) In 3.7 we saw that there were grounds for a general objection to the Identity Theory taking the form of a demand for an explanation: to explain why what seems to be additional to neural events is not additional.

The I-Identity Theory was troubled here because it maintained that mental events were only in fact not additional to physical events, but there was no factual explanation available. (The sort of explanation envisaged in an internally stable account would involve undermining the dualist's resistance - which would in turn involve undermining the objections listed above.)

(v) We may add, for good measure, an objection based on the Part 2 notion of Residual Privacy (cf. 2.3).

A dualist may argue that a person is able to make an introspective distinction between isomorphic visual imagings (or perceptual events generally) - a distinction which is in principle incommunicable by him to others. Thus it appears that someone else could not determine which of two isomorphic visual imagings was occurring in a person, though he could in

principle determine which ϕ (identity candidate) was occurring.

I will now in Part 4 attempt to undermine the coherence of the supposition that (o) requires (p) - I will try to undermine the intuitive basis for dualism. The above objections are merely fragments of a dualist view, fragments which will hopefully be disposed of along with the view which nurtures them.

NOTES TO PART 3

[1] Though not stated explicitly, this is at least suggested by Feigl's:

> Utilising Frege's distinction between Sinn... and Bedeutung..., we may say that neurophysiological terms and the corresponding phenomenal terms, though widely differing in sense..., do have identical referents. ((Feigl (2), p. 38.))

It is questionable to what extent Sinn and connotation match. But in the Identity Theory context they were taken to be roughly the same.

[2] Smart considers this kind of thought-experiment presented as an objection to his view: 'I can imagine myself turned to stone and yet having images, aches, pains, and so on.' ((1), p. 63). His defence rests on the logical possibility of dualism: the dualist

> holds that experiences are something over and above material processes, that is, that they are a sort of ghost stuff... I say that the dualist's hypothesis is a perfectly intelligible one. But I say that experiences are not to be identified with ghost stuff but with brain stuff. This is another hypothesis, and in my view a very plausible one. (((1), p. 64.))

[3] Kripke is illuminating on some of the general issues (i.e. not specifically mind-body issues) which are involved in the discussion above. But ironically, where he specifically considers the mind-body issue, he implicitly supposes there are two phenomena involved (as we shall see later).

[4] Showing that (p) is not required by (o) is not enough to show that mental events must be physical. To show the latter it must be shown that (p) cannot be coherently derived from (o). We must remember (cf. 2.61) that the PLA does not show that the derivation of (p) from (o) is not coherent - it only shows that "(o) and (p)" is not a coherent conjunction.

Generally in Part 3 I speak of the Identity Theorist's duty merely to show that (o) doesn't require (p) - although this would not match the consequence of the PLA, that (p) is impossible. In the light of the PLA we must expect a satisfactory account to show that (o) can't coherently require (p). My justification for using the weaker '(o) doesn't require (p)' is, as I will mention shortly, that I will generally put in abeyance (in Part 3) the impact of the PLA - that dualism is logically impossible.

5 Hinton, however, has an obstruction to the Identity Hypothesis relevant to this point. Hinton's obstruction is based on the view that, at least in certain crucial contexts, expressions purporting to refer to events are not to be regarded as genuine referring expressions. Hinton traces his point to Geach, who suggests that:

> we need to get events expressed in a propositional style, rather than by using name-like phrases (what Kotarbinsky has called "onomatoids"). We need, that is to say, propositions like "Wellington fought Napoleon at Waterloo after George III first went mad", rather than "George III's first attack of madness is earlier than the Battle of Waterloo". ((Geach, p. 313.))

Geach's ontological preference for people and things is supported by the reductive eliminability of event-phrases:

> "The news of Queen Anne's death made Lord Bolingbroke swear" goes over into "Lord Bolingbroke swore because he heard Queen Anne had died". Cutting out the onomatoids in this way, we get a manner of speaking in which persons and things are mentioned but events do not even appear to be mentioned; ((Ibid., p. 313.))

Behind this grammatical elimination lies the philosophical consideration that if there are people and things, and that is all the world contains, we just don't need events; indeed, we couldn't have the event: Queen Anne's coughing, and the person: Queen Anne, who coughed, or there would be too many things in the world.

Hinton is thus able to argue as follows (cf. Hinton, p. 244). (I convey just the gist of his argument, which he applies to a statement of the economy argument for the Identity Theory. But it applies equally well to an attempt at stating the Identity Hypothesis.)

Suppose that ϕ is the neural change with which Bill's phosphene (an illusion of a flash of light) is to identified; then the Identity Hypothesis would go like this:

(1) ϕ at time t was an actual event;

(2) Bill's phosphene at time t was an actual event;

(3) One and only one actual event has been mentioned.

But if there is a reductive analysis of the predicate '... is an event', the first two premises reduce to:

(1r) Bill "ϕ-ed" at time t;

(2r) Bill "phosphened" at time t;

(where verbs are introduced to say that Bill underwent the relevant neural change and that he had a phosphene).

Now no actual events are mentioned - only Bill is mentioned. In other words, if the world contains just people and things, it would not be sensible to have event-identity.

However, this obstruction to the Identity Hypothesis does not appear to me to be successful.

Hinton makes the assumption that the logical subject in (1r) and (2r) is one and the same. Certainly in (1r) and (2r) only Bill mentioned. But it is not clear that a dualist would allow (1) and (2) to reduce this way. For he would want subject terms with different connotations to occur in (1r) and (2r), for instance, as in:

(1r') Bill's body ϕ-ed at time t;

(2r') Bill (the person or subject of consciousness "associated with Bill's body) phosphened at time t;

and then, the dualist would say, two things are mentioned, namely Bill and Bill's body. And the Identity Theorist would say that only one thing was mentioned. This however is not entirely satisfactory for the Identity Theorist, since he wishes to say more than just that the logical subjects in (1r') and (2r') have the same referent. For in

(4) Mr. Hyde scratched his ear at time t;

(5) Dr. Jeckyll made out a prescription at time t;

there is also one referent for the two subject terms. But one would want to say, in the ordinary event-idiom, that (4) and (5) describe two events; whereas the Identity Theorist would want to get across the significance of the ordinary idiom, that (1r') and (2r') describe one and the same event - that ϕ-ing is phosphening.

But I think Hinton himself provides a clue as to how this could be

achieved. For the point of a reductive elimination of the sort Hinton engages in is not to restrict our powers of expression but to translate or convert our means of expression so as to reveal a more economical ontology. Hinton says that event-statements 'may be said to "mention an event", in some sense of the phrase' (ibid., p. 244). And this is to be taken to mean that each event-statement is a syntactically proper answer to the question 'What happened at t?'. Thus he says:

> 'The prince was turned into a toad', and so forth, belong in the syntactical box or category labelled 'Statements that answer the question "What happened?" ' ... It is a logico-grammatical fact that statements like 'The prince was turned into a toad' answer that question; and this logico-grammatical fact is surely the only fact I stated when I said: '... a prince's being turned into a toad is a happening' - or '... is an event'. ((Ibid., p. 245.))

We say: 'the beheading of Charles I and his execution were the same event', or: 'the boiling of the water and the vapour pressure of the water reaching atmospheric pressure are the same event'. And we must still be able to get these across even if there are no events. (Indeed, a test of a successful reductive elimination is precisely whether we can get such statements across without 'event' terminology.) One way we could do this, perhaps, is by saying of, for instance

(6) The water in the beaker started to boil at time t;

(7) The water-in-the-beaker's vapour pressure reached atmospheric pressure at time t;

that there is a particular question of the form 'What happened at t to the water?', to which (6) and (7) are both equally correct answers. Another less problematic way would be by following a suggestion of Hinton's:

> We can, surely, say that the beheading of Charles I and his execution were the same event, because that was how he was executed. But it would be strange to admit, on this ground, that the class of events has at least one member, if we are unwilling to admit it on the ground that Charles I was executed. ((Ibid., p. 248.))

But if a reductive elimination of 'events' is successful, it shouldn't, in a sense, matter: we should still be able to say all we want to say; (except of course for 'there are actually events in the world'). And there is no

reason, therefore (at least, on the ground that '... is an event' can be eliminatively reduced) why the mind-brain event-identity should be worse off than any other event-identity. So we should be able to say something like 'Bill phosphened by ϕ-ing' (having established that there is only one Bill - Bill's body). (It may be argued, of course, that whereas 'Charles I was beheaded' could be an explanation of how Charles I was executed, 'Bill ϕ-ed' could not be an explanation of how Bill phosphened. But perhaps this disparity assumes <u>other</u> (more conventional) objections to the Identity Theory.) In other words, there is no reason to suppose that the hypothesis that Bill's ϕ-ing <u>is</u> his phosphening should be seriously obstructed by a successful reductive elimination of the noun phrases, unless that reason is something additional to the eliminability of events (which programme should be neutral as regards dualism or materialism). We should still be able to say: 'Bill didn't ϕ <u>and</u> phosphene, he just ϕ-ed'.

6 My argument is based on Hinton's (cf. Hinton, pp. 251 - 252). Essentially, Hinton points out that an <u>adequate</u> meaning-analysis of a sensation report would by its intensionality preclude substitution of 'ϕ' in the analysis <u>salva veritate</u>.

7 Even Hinton's improvement on a meaning-analysis is not sharp enough, for "something which I find like..." could be a feeling of nausea in the example where seeing blood was the stimulus condition - one would have to focus on the respect in which one "found it like...".

8 We must expect this in a view which does not satisfactorily show that (o) doesn't require (p) - which does not show that we can be actually introspectively aware of mental events which have physical (neurophysiological) descriptions. For then if (p) is denied, and '(o) requires (p)' is implicitly allowed to stand, then (o) is implicitly disallowed.

9 This problem has many alternative voicings - for instance Nagel's:

> It is useless to base the defense of materialism on any analysis of mental phenomena that fails to deal explicitly with their subjective character. For there is no reason to suppose that a reduction which seems plausible when no attempt is made to account for consciousness can be extended to include consciousness. ((Nagel (3), p. 437.))

[10] Armstrong also wants to hold a view which retains (o). Thus he says, criticising Ryle's behaviourism:

> When I think, but my thoughts do not issue in any action, it seems to me as obvious as anything is obvious that there is something actually going on in me which constitutes my thought. ((Armstrong, p.72.))

[11] Smart has another counter to Baier, however, in which he tries to oppose directly Baier's contention:

> Both the materialist and the dualist, and indeed most behaviourists also, will want to say that the sincere reporting of a sensation is one thing and the sensation reported is another thing. Now, as Hume said, what is distinguishable is separable. It is therefore logically possible that someone should sincerely report an experience and yet that the experience should not occur. ((Smart (3), p. 108.))

This is a much more fruitful counter. Note that whatever the efficacy of this counter, it does not rely on the truth of the Identity Theory for its efficacy: it is independent of the Identity Theory, and tries to undermine the coherence of the view that the relevant reports <u>are</u> incorrigible. Thus this reply tries to undermine the intuitive resistance to the Identity Theory, rather than suppose that this resistance does not have to be undermined since it cannot damage the Identity Hypothesis. I will revive this sort of counter in Part 4.

[12] I will comment later on the possible qualification that the judgement should be sincere and linguistically competent. It might be objected that even allowing for this qualification, ~I/. Austin for instance gives an example suggesting this objection:

> ... 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 <u>also</u> that I may be brought to see, or perhaps remember, that the colour before me just wasn't <u>magenta</u>. And this holds for the case in which I say, 'It seems, to me personally, here and now, as if I were seeing something magenta', just as much as for the case in which I say, 'That isn't magenta'. The first formula may be more cautious, but it isn't <u>incorrigible</u>. ((Austin, p. 113.))

It seems though that the subject may still be allowed final epistemological authority: for it could be argued perhaps that it could not be established that it did <u>not</u> seem to the subject <u>as if</u> he were seeing magenta until the subject himself admitted his mistake. The possibility of the subject

"changing his mind" is most plausible in the case of inattention on his part. At any rate, I will assume here, as Bradley does, that $I\not\alpha$.

[13] Bradley also considers an objection based on the incorrigibility predicate divested of the modal operator. It might be supposed that such a predicate has not the intensionality of I which appears to obstruct the incorrigibility objection. Bradley takes this new predicate, which I will call 'K', such that 'Kx' is read as '~(x is judged to occur and x does not occur)', (ibid., p. 177). The modified incorrigibility objection would use the necessity that, by Leibniz's Principle, $(x = y) \supset (Kx \supset Ky)$, and attempt to show that $K\not\alpha .\ \sim K\phi$.

Bradley then makes two points:

(a) every occurrence of 'x' in 'Kx' which is to be replaced by 'y' on application of Leibniz's Principle must be purely referential - in particular, 'x is judged to occur' must be a referentially transparent context such that $(x = y) \supset (x$ is judged to occur $\equiv$ y is judged to occur); otherwise it would be possible that $x = y.Kx.\ \sim Ky$.

(b) whereas $\sim\Diamond I\phi, \Diamond K\phi$: it is always possible that ~(φ is judged to occur and φ does not occur); and since it is necessary for this objection that $(\not\alpha = \phi) \supset (\not\alpha$ is judged to occur $\equiv \phi$ is judged to occur), it can only be contrived that $K\not\alpha .\ \sim K\phi$ on the assumption that $\not\alpha \neq \phi$ - but then the Identity Hypothesis is contingent; so this is no objection against the possibility that $K\not\alpha .K\phi .\not\alpha = \phi$.

So, unless 'judged' is taken as providing an intensional context (which would of course be the usual idiom), in which case Leibniz's Principle wouldn't apply anyway, the fact that $\Diamond K\phi$ means that K cannot provide an effective objection to the Identity Hypothesis.

[14] Quinton, for instance, traces this difficulty to the epistemic element in the notion of incorrigibility (cf. Quinton, p. 348). And this counter

is also related to a general misgiving concerning the incorrigibility objection, that incorrigibility is primarily a property of certain reports rather than a genuine property of ϕ.

[15] It was pointed out to me by Mr. P. F. Snowdon that care must be taken in construing 'A' occurring inside the quotation marks in the predicate construction. For if 'A' is construed as a description, A could be ignorant or wrong about the reference of 'A' without being linguistically incompetent - in which case it may be possible that $\sim I'_{\psi}\ \phi$. But we may stipulate that 'A' be a proper name, in which case we can regard ignorance of its reference as linguistic ignorance; or at any rate we may assume that A knows how to use correctly the name 'A'.

[16] A view such as that of W. and M. Kneale precluding a de re interpretation of 'x is under any description necessarily F' (cf. Kneale, p. 618) would go against the incorrigibility objector - but there is considerable opposition to the Kneales' view.

[17] This point is essentially one made by Bradley, in his case applying to the question of the existence of some manner of detection by sensory methods of something at a certain spatial location. He points out that to infer the occurrence of sensations from certain observed neural conditions would constitute a manner of detection (cf. Bradley, p. 182 - 183). In the case under consideration, "perceiving that", we wouldn't even need a rigorous theory: ordinary everyday bases for mental ascriptions would do.

[18] The problem of volition, and of the interaction, if any, between mental and physical events, is largely taken up within the Identity Theory controversy by the location issue, since naturally an interaction between a located and an unlocated event is difficult to conceive.

[19] There is a less plausible location objection, which maintains that ϕ is spatially located, but not where ϕ is. For instance, it is admitted that it does make sense to say of at least some somatic sensations that they're located in a region of space: thus, a pain in someone's foot is

admitted to be occurring <u>in</u> his foot. But then the pain would not be in the same place as the most plausible identity candidate ϕ, since ϕ would be some process of the central nervous system, and not any process occurring just at the place of the pain. (Cf. Bradley, pp. 181 - 189 , who advocates this form of the location objection.)

But there are various considerations which suggest that a locution apparently locating a pain in a certain part of the body should not be taken as implying that any mental event actually occurs in that part of the body. The most obvious support for this comes from considering the phantom limb case, which provides an example where a ϕ could be identified as introspectively indistinguishable from a pain-in-A's-foot, but could not possibly be occurring in A's (amputated) foot. Which suggests that the locution 'in-A's-foot' should be regarded as a means of individuating a certain mental event, rather than actually spatially <u>placing</u> it. Perhaps the locution 'a hurting of my foot', like 'a memory of Huddersfield', is less disposed to suggest that the hurting is located spatially at the place used in the individuation. (Brentano used the phantom limb example to illustrate this sort of point (cf. Brentano, p. 85).)

[20] This sort of analogy is central to the Local-Sign Theory, and is made by James (ibid., p. 158).

[21] Clearly this sort of view is related to a Stage I view in which the relations on <u>private</u> stimuli are maintained to be distinct from those relevant to a description of the external environment. The sort of location objection set in such a view is a reaffirmation of the intuitive pressure for a Stage I view.

[22] There are complications to an example like this which I will mention in Part 5.

[23] This would not really alleviate the problem that a person seems to be able to be introspectively aware of <u>some</u> change occurring in him when he has a yellowy-orange after-image - that event just doesn't seem like the

neural event he might imagine being perceptually aware of.

[24] Spatial contiguity could not be used as a means of separating the events by their contigous causes and effects, since the principle of the spatial contiguity of cause and effect is 'an inevitable casualty of the dualist doctrine' (Quinton, p. 336).

[25] Nagel defends the Identity Theory against the location objection by a device which may be suspected of containing a suppression rather like Quinton's:

> Instead of identifying thoughts, sensations, after-images, and so forth with brain processes, I propose to identify a person's having the sensation with his body's being in a physical state or undergoing a physical process. Notice that both terms of this identity are of the same logical type, namely (to put it in neutral terminology) a subject's possessing a certain attribute. ((Nagel (1), p. 216.))

Like Smart's proposal, that the referent of 'having a sensation' rather than 'sensation' should be sought (e.g. Smart (4), p. 161), Nagel's move appears to eliminate the problem of finding a physical process which a certain sensation must be. But, in addition, Nagel is able to argue that:

> ... we can be rid of the stubbornest objection ..., that having to do with location. Brain processes are located in the brain, but a pain may be located in the shin and a thought has no location at all. But if the two sides of the identity are not a sensation and a brain process, but my having a certain sensation or thought and my body's being in a certain physical state, then they both will be going on in the same place - namely, wherever I (and my body) happen to be. ((Ibid., pp. 217 - 218.))

But a dualist would argue that Nagel is suppressing an essential feature of the location objection: the dualist will suggest that it is not at all clear that the subject of mental experience is in the same place as, indeed, is spatially related at all to, the body "associated" with that mental subject. In other words, it will be argued that it is not a person's body which is referred to by the subject terms on both sides of Nagel's Identity Hypothesis, for it is not the person's body which has the sensations, but the mental subject "associated" with the body.

I don't want to support the dualist's argument: - I want to emphasise that the Identity Theory cannot afford to leave unopposed such intuitive

views as the dualist's here. The strategy of Nagel (and others) is clearly one of hoping to avoid a direct confrontation with such a view. But this confrontation cannot be avoided.

[26] Indeed, it may be argued, for instance, that every individual wedding is a physical event. But we could not locate the weddings independently of the physical events with which they are to be identified.

[27] Of course, to achieve (o) without (p) successfully would entail that "blank automatism" itself is incoherent.

[28] In Part 4 I will suggest an account in which we can avoid supposing that when a person feels pain he is ipso facto in an epistemic relation to the pain; but that in a sense in which we could allow that a person was introspectively aware of a feeling of pain, for him to be introspectively aware of a feeling of pain is not for him to feel pain. But I will not return to discuss Kripke's argument explicitly.

[29] This may seem odd, but because we expect that the purpose of designing such a machine is geared to that background. But a scientist from a different cultural background might discover such a machine, and though he may be able to explain its behaviour electronically, he may have no idea of the interpretation available: of what the machine was doing in a chess game. He may be able to guess, of course - to reconstruct the game of chess - but he needn't have to.

[30] Putnam also has arguments suggesting that a dualism of descriptions is applicable to purely physical structures (cf. Putnam (1)). Putnam's conceptual dualism differs from Davidson's, in that his involves a distinction between the "logical" and the physical state of a physically realised Turing machine: a physical state of the machine may be described in intrinsic physical terms, or as a logical state. (Thus: 'I am in a state A if and only if flip-flop 36 is on' ((1), p. 74).) The distinction between a logical state and a physical state involved here is roughly the sort of distinction involved in the contrast between an abstract position on an

abstract chess board, and a physical set up on a physical board instantiating the abstract structure. In Putnam's case there is a strict numerical correspondence between the two sorts of description, unlike Davidson's much looser dualism of description. But again as in Davidson's account, Putnam's conceptual dualism does not suggest ontological dualism.

31 Thus Wundt, who held a parallelist view, says:

> ... an external voluntary movement is not produced by the internal act of will, but by the cerebral processes correlated with it; an idea does not follow from the physiological excitation of the sensory centre, but from the processes... which run parallel to them. ((Wundt, pp. 449 - 450.))

32 A remark of Spinoza's seems relevant here:

> No one yet knows in what way and by what means the soul moves the body, nor how many degrees of motion it can give to it, nor with what rapidity it can move it. Whence it follows that when men say that this or that action is produced by the soul which has dominion over the body, they really do not know what they are talking about, and are only confessing in terms flattering to their vanity, that they are ignorant of the true cause of the bodily actions which surprise and astonish them. ((Spinoza, pp. 56 - 57.))

This remark suggests the possibility that we have to presume that the true cause is a bodily process, as Davidson does. But ontolcgical monism does not follow just from this.

33 Davidson's general point, construed as a form of the Materialist Compatibility Point, suggests that arguments such as (one of) Malcolm's (cf. Malcolm, pp. 176 - 178) maintaining that a conceptual dualism in virtue of being a <u>conceptual</u> dualism is damaging to physicalism, cannot be taken as troubling physicalism in its weak form. (In Part 5 I will discuss again an example involving a chess machine, and suggest that we can extend Davidson's view to allow ontological monism as sufficient for the strong form of physicalism.)

Also, if Davidson is right about the irreducibility of mental discourse to a physicalist discourse, a "disappearance" form of the Identity Theory would not be feasible. A "disappearance" view such as Rorty's may be construed as a view which denies (o) and (p), but which suggests that the

linguistic disappearance of '$\not\phi$' is the solution to the semantic analysis of '$\not\phi$' (though we may retain mental discourse for convenience): in other words, in principle we could use 'ϕ' where we at present use '$\not\phi$'. So it is not really maintained that '$\not\phi = \phi$':

> ... the relation in question is not strict identity, but rather the sort of relation that obtains between, to put it crudely, existent entities and non-existent entities when reference to the latter once served (some of) the purposes presently served by reference to the former - the sort of relation that holds, for example, between 'quantity of caloric fluid' and 'mean kinetic energy of molecules'. ((Rorty, p. 189.))

PART 4

SUBJECTIVE DUALISM AND ONTOLOGICAL MONISM

I will now present an account in which the sort of considerations which the dualist supposes favour ontological dualism can be seen as compatible with ontological monism. I will suggest that there is a view that can be justified, which I will call subjective dualism; but that the dualist illegitimately and in confusion takes the justification of this view, which is compatible with ontological monism, as a justification for ontological dualism.

I will briefly and loosely reconsider the intuitive background to ontological dualism; and then I will give an informal account which tries to illustrate the sort of confusion the dualist is guilty of. This informal account contains some misleading simplifications in order to facilitate the presentation:- it should be regarded only as a crude way of illustrating the dualist's confusion. Moreover, the dualist account I expose as confused is very intuitive and unsophisticated - and it may not be the intuitive picture with which every ontological dualist would care to be associated. I hope though that it does represent roughly the background against which the Identity Theory seems implausible.

I will follow this illustration by a rather formal account using elementary set-theoretical concepts and notation. The application of this formal account to mental events is intended to give substance to the notion of subjective dualism and its compatibility with ontological monism; and also to render more accurately the confusion indicated in the preliminary illustration. And to conclude Part 4 I will indicate how we may accordingly supplement the defence of the Identity Theory.

4.1 Monism Resisted

Let's suppose that someone is led to an ontological dualism of mental and physical events through the following considerations. (I will concentrate on

visual matters for simplicity.)

This person accepts that when, say, he sees a cat jump off a chair, there occur the three events: the cat jumping off the chair, a certain neural event occurring in him, and his visual imaging "of", associated with, "suggesting" the event involving the cat. The third event - the mental event - may not have to be describable as a visual imaging; but describing it as such tends to strengthen the intuitive contrast between it and some appropriate neural event - a contrast which it is my aim to undermine. I will speak of the visual imaging "suggesting" the environmental event simply to bring out the point that the visual imaging is supposed somehow to convey information to the "subject" about the environmental event. And we may take it that our dualist allows that a visual imaging may occur in him without the "suggested" event occurring - he would probably describe such an event in him as a visual illuding. Although I will not discuss visual illudings or visual events, such as after-imagings, which would not be taken as "suggesting" any environmental event, such events will be accommodated by the account I develop.

Suppose our dualist also accepts that events "similar" to his visual imagings occur in other people. And he notes with regard to his visual imagings:-

(a) He does not <u>see</u> his visual imaging - for example, one "suggesting" a cat jumping off a chair - in the same sense that he sees a cat jumping off a chair (or in any sense at all).

(b) And he does not see anything happening in another person and believes that no examination of another person's body, however detailed, will reveal anything "like" his visual imagings. Though such examination might reveal events like his neural events.

(c) He concludes that a person's visual imagings are in principle imperceptible to others:- however much he may discover indirectly about someone else's imaging, he could never perceive that imaging; though the events he can perceive (physical events) may be evidence of some sort for the occurrence of imagings. And he concludes that there are two sorts of events, those that

in principle could be perceived through some sense modality (physical events) and those which in principle could not, and to which only one person has "direct" access (mental events).[1]

This conclusion is highlighted by a thought-experiment involving auto-cerebroscopy. Suppose the dualist is scanning his own brain (to any desired degree of detail), and perceives, or indirectly deduces as occurring through the theoretical equivalent of a microscope, a certain sequence of events which seem to be correlated with the sequence of visual imagings occurring as he scans his own brain. Wittgenstein mentions this sort of example:

> ...the subject is at the same time the experimenter, who is looking at his own brain, say by means of a mirror. (The crudity of this description in no way reduces the force of the argument.)
> Then I ask you, is the subject-experimenter observing one thing or two things?...The subject-experimenter is observing a correlation of two phenomena. One of them he, perhaps, calls the <u>thought</u>. This may consist of a train of images, organic sensations, or on the other hand a train of the various visual, tactual and muscular experiences which he has in writing or speaking a sentence. - The other experience is one of seeing his brain work.
>
> ((Wittgenstein (1), p.8.))[2]

Now, someone else scanning our dualist's brain at the same time would perceive or deduce as occurring a sequence of cerebral events, and could not, according to the dualist, be perceiving the dualist's visual imagings. Moreover, the dualist would surmise that his co-scanner would be having his own visual imagings, but would be <u>seeing</u> the dualist's cerebral events. Both the dualist and his co-scanner would be seeing the same sequence of events - perhaps (picturesquely), things happening to grey matter - but they couldn't both be seeing the dualist's visual imagings - indeed, the dualist is not seeing them, though he is "having" them.

So again the conclusion of our dualist is that visual imagings could not be cerebral events - though these <u>two</u> sorts of events may be correlated.

The considerations which lead our dualist to his conclusions are naive, perhaps, but compelling; and something like them forms the intuitive basis for most arguments against ontological monism. However, by furnishing the dualist's points with the concept of the contrast between the syntax and the interpretation of a language, we can roughly illustrate his points while

absolving them from a dualistic bias.

4.2 The Resistance Undermined

Let's suppose that someone's only visual contact with his environment has been through messages which he receives in some medium; the messages are of the form 'X occurs', where 'X' is an appropriate visual description of the events which this person, S, scans.

The points I will make through this example should be seen as illustrations - one important feature of the example which does not accurately reflect the situation to be illustrated is the retention of the "subject" who apprehends the messages; it is easier to introduce the essence of the dualist's confusion by making this simplification.

I will now consider some of the points which occur or are implicit in our dualist's "argument", and comment in turn.

(i)' A visual imaging is "associated with" or "suggests" an event taking place in the subject's environment.'

This "association" is to be rendered in the present example through the analogy of the association between a description contained in one of S's messages and what it describes. For simplicity we will take such descriptions to be definite, the analogous association then being the relation between a referring expression 'X' contained in a message, and its referent.

(ii)'One does not <u>see</u> one's visual imaging "suggesting" a cat jumping off a chair when one sees a cat jumping off a chair - one "has" the visual imaging'.

Analogously, in the illustration we find that there is going to have to be a distinction between the sense in which S "apprehends" the message 'X occurs', and the sense in which S "apprehends" X. The first I will call 1-apprehension, the second 2-apprehension. Perhaps a slightly different example might help to make this distinction clearer. Suppose that a vehicle is provided with such sensitive radar equipment that it can be driven with its windows blacked out. The driver "apprehends" the radar screen in a different sense from that with which we may say that he "apprehends" the environment outside the vehicle

by means of the radar screen. I am not here concerned with a particular problem of conceptual/verbal classification - with the problem of whether a certain process falls under the concept "seeing" - I'm concerned with a general basis for the division of certain pairs of concepts. "Apprehending" or obtaining information about an object, and "apprehending" or obtaining information about something else by means of that object, gives the genus of these pairs of concepts. Of course, in the example of the driver, he doesn't have to apprehend the extra-vehicular environment by means of the radar screen - he can see both the screen and the road (if he clears his windscreen); information about his environment is available not only through the radar screen. But the particular species of this genus of pairs with which I wish to illustrate the case of visual imagings is where information about an external environment is only available through some medium. This is to be the case in our example with S (we could have stipulated this in an example of the radar screen variety).

There is of course one very important difference between these illustrations and the case of visual imagings: in the examples we clearly have a subject/object picture of 1-apprehension. We know that this will not work for visual imagings. But this does not matter at the moment: it will not hurt my argument that the sort of confusion to be illustrated through these examples is not properly rendered by them. For (a) even with an erroneous subject/object picture, ontological dualism will not be sustained - as we shall see shortly; and (b) the dualist's confusion which will be more accurately described later on is still, I think, usefully anticipated in this way. We may for the moment allow some parallel between the dualist's "awareness" of his visual imagings and 1-apprehension; and between his awareness of his visual environment and 2-apprehension.

(iii) 'No examination of the world (including of course other people) can reveal a visual imaging.'

If we transpose this contention to the example with S we can begin to see how certain relevant confusions may arise. The parallel supposition by S is that he could not 2-apprehend the occurrence of a message. A simple confusion

might lead S to this belief:-

Suppose S 1-apprehends the occurrence of the message M, and considers whether he could 2-apprehend its occurrence. S may suppose that if he is 2-apprehending the occurrence of M then he should be able to compare the object of 2-apprehension, the occurrence of M, with an object of 1-apprehension, also the occurrence of M, and find that they are the same. This seems reasonable enough. But suppose S 2-apprehends the occurrence of M through the message M'. S might in confusion compare M and M', find that they are different, and conclude that he is not 2-apprehending the occurrence of M (through M'). Following this mistaken proceedure, S might even come to believe that he could never 2-apprehend the occurrence of a message. We can also show this confusion by means of the radar screen example. It is quite possible in principle that the driver should detect on his radar screen the occurrence of a picture on a radar screen in another truck - as the occurrence of a certain configuration in his picture, and, finding them to be quite different, incorrectly deduces that he is not detecting the occurrence of something like what is happening on his screen.

I will say more about this sort of confusion presently.

Let's now consider the conclusion of the dualist's "argument":

(iva)'There are two sorts of events: those that can or could in principle be perceived through a sense modality - those "sugrested" by imagings - and those which <u>are</u> imagings, and which could not be perceived.'

(ivb)'Supposing an imaging ψ was supposed to <u>be</u> a cerebral event ϕ. But if we could see the cerebral event ϕ we wouldn't see the imaging ψ; and if we could see (autocerebroscopically) the cerebral event which was supposed to <u>be</u> the visual imaging "suggesting" <u>that</u> cerebral event, there would surely be two things, the visual imaging, and the cerebral event seen.'

In our example (iva) would transpose to the contention that there would have to be two sorts of things, X's, and the messages of the form 'X occurs'. And (ivb) would transpose to the contention that there would always have to be two things, X and 'X occurs'.

But (a) there is a case where there would not be two sorts of things, the messages - syntactical entities of the form 'X occurs' - and the X's - referents of the expressions contained in the messages - simply: where the messages are among the things referrable to by the expressions contained in the messages. We may call this the case where syntactical reference is possible.

And (b) there is the case where there may not even be two things, X and 'X occurs', but one. In this case the message M is self-referential, in that the expression 'X' in M refers to M. This then is the case where self-reference is possible.

But this is a rather special case: when we come to consider the notions of syntactical reference and self-reference applied to visual imagings, we will find that the self-referential case cannot arise. But the possibility of syntactical reference alone is sufficient to collapse any dualist bias in the analogy; and it is essentially this possibility of syntactical reference, applied to visual imagings, which will do the work of collapsing dualism.

For our analogy supports dualism only if dualism is built into it by excluding the possibility of syntactical reference. If we modify the analogy so that we explicitly allow syntactical reference, and take the analogy back to illustrate the imaging/cerebral event case, we can correspondingly remove the bias towards the imaging/cerebral event dualism.

Thus if we take the analogy back, there remain the two events: the visual imaging "suggesting" a cat jumping off a chair, and the event seen - a cat jumping off a chair -(in the analogy: there is 'X occurs' and X); but we allow that the imaging might be <u>seen</u> (or indirectly perceived or deduced to occur, as are many physical events). In the analogy, the possibility of syntactical reference leads to the possibility that S may 2-apprehend one of his own messages, or someone else's message (supposing others to be in S's predicament): when he 1-apprehends 'X occurs' and 'X' refers to a <u>message</u>. So now we allow that there might be an imaging which "suggests" an <u>imaging</u>:- for example, a visual imaging "suggesting" a visual imaging "suggesting" a cat jumping off a

chair. And if an imaging is not visible in this way, this need not be taken as indicating any qualitative difference in ontological status between an imaging and any other event which is (contingently) not visible.

So if the equivalent of syntactical reference is possible in the case of visual imagings, an examination of the world may (visually) reveal to our dualist a visual imaging - in just the way in which he supposes the world is revealed to him. The confusion mentioned with regard to (iii) above, by which S might be led to suppose that he could not 2-apprehend a message, is paralleled by the following confusion in the dualist's line: having distingusished between an imaging and the event "suggested" by it, he supposes that he could not see a visual imaging because he confusedly compares (or imagines himself in a position to compare) a visual imaging "suggesting" a neural event ϕ with the visual imaging ψ which that neural event is supposed to be. And finding the two imagings to be different, he mistakenly deduces that the neural event ϕ cannot be the visual imaging ψ. In other words, the dualist makes a conceptual distinction without which an imaging/neural event dualism, based on the imperceptability of imagings, cannot get started:- the distinction between an imaging and the event it "suggests" (- a realist distinction). But he allows this distinction to lapse when he confusedly fails to distinguish between the imaging "suggesting" ϕ, and ϕ.

But it is not really surprising that this sort of confusion occurs: we naturally become so "familiar" with imagings that though we may be able to make the conceptual distinction between an imaging and the environmental event "suggested" by it, awareness of this distinction is not needed in order to react effectively to environmental events. Indeed, most organisms cannot make this conceptual distinction between environmental events and events which occur in them which are in some manner isomorphic to the environmental events. We may understand by the "familiarity" with imagings which I mentioned just now simply the idea that we can and normally do react to the environment on the basis of perceptual events without making this distinction; and without

the occurrence in us of mental events describable, perhaps, as "noticing or thinking about perceptual events occurring in us". We are normally as unaware of the imaging/environmental event distinction as we are unaware of the numeral/number distinction when we perform arithmetical calculations. So it becomes very easy to overlook the "designatory" role of the imaging. And then what seems so odd is that a neural event could be an imaging ϕ, or the experiencing of a pain: because we try to compare the imaging "suggesting" the neural event with the imaging ϕ, or the experiencing of pain. It is as though we made the elementary mistake of confusing 'X' with X, and thus found it hard to accept that X=Y because a comparison of 'X' with Y suggests otherwise.

But if we can accept the hypothesis that among the events which could be "suggested" by imagings are the imagings themselves, then the resistance to monism may be undermined. If we can apply the syntax/interpretation distinction to the case of mental perceptual events (as I will in 4.4), the monistic hypothesis becomes the hypothesis that syntactical reference is possible in that case:- in the account I give in 4.3 and 4.4, we will see how there may be just physical events, though certain interactions between physical events may be regarded in such a way that one event can be interpreted as a designation, another as a designatum. And we will have available a more precise sense for speaking of an imaging's "suggesting" an environmental event: the designatory role of imagings will become clearer.

There is another aspect of the view I will present, which I will call subjective dualism. Essentially this will claim that the event designated by a mental perceptual event must be distinct from that (designating) mental perceptual event. So if an imaging were to designate an imaging, the former would have to be different from the latter. But if this difference is taken as a difference between an imaging and a neural event (as I will suggest it is sometimes taken in thought-experiments involving the comparison of mental events and neural events), then subjective dualism can mistakenly lead to ontological dualism.

By way of a transition from the illustration we have been considering to a more accurate account by which we can render the designatory role of imagings,

I will emphasise the following points.

In the view I will endorse we will make a realist distinction between an event occurring in what may be described as an observer (or subject), and the observed event (normally in the observer's environment) which is designated in a special way by the event in the observer. The observed events may also be designated or described, in the normal linguistic manner, when we describe in natural language the event which the observer perceives. For example, 'The explosion in Tycho's Star' designates an event which occurred years before the appropriate visual imaging in Tycho Brahe; but the latter event designates the former event in the special way I will describe.

But often we find, attending a realist distinction, the introduction of the notion of sense-data which <u>mediate</u> between a person and his environment. But a mediatory view of this sort must be abandoned - indeed, it is primarily in this respect that we must part company with the illustration involving S and his messages. For to suppose that a person "1-apprehends" his sensory imagings, and interprets these so that he can "2-apprehend" his environment is to make the mistake exposed by Wittgenstein and Ryle - of supposing that a person stands in relation to his sensory imagings in a manner analogous with his relation to the external objects perceived. We must dissociate ourselves from a view in which a person is aware of his mental events as they occur - even if such awareness is maintained in a different sense ("1-aware",perhaps) from that with which the person is said to be aware of what he perceives. So we must dispense with the notion of a subject who "1-apprehends" his visual imagings:- the latter <u>just occur</u>. (Of course, we can still allow talk of a subject <u>having</u> visual imagings, of visual imagings occurring <u>in</u> a subject, and so on - provided we don't construe this as implying an awareness relation between the subject and his imagings.) I must stress, though, that this does not preclude there being a sense in which we <u>can</u> be aware of our visual imagings: roughly, in that we can allow that events occur in us which are introspective. But it must not be supposed that because there <u>can</u> occur events in us describable as thinking about,noticing,paying particular attention to events <u>occurring in us</u>, such events occur when we normally perceive things, feel pain,

and so on. Similarly, when such a train of thought occurs in a person, it does not have to be "examined" or noticed - it just occurs. When we think we don't think and notice our thoughts, although we can, in the middle of a train of thought, think 'Ah! A train of thought is occurring in me' - that sort of event is also possible. And when that sort of event occurs, we can be said to be conscious of our thoughts.

In this way we can dispense with the notion of a ghostly subject who is aware of his mental events; but without precluding the possibility of introspection - without precluding the possibility of a genuine sense in which a person can be aware of his mental events. But this latter possibility boils down to the possibility, not of there being a subject who is aware of his imaging, but of there being two events: for example, an imaging and a train of thought about the imaging.

Thus, suppose a visual imaging could be and was induced in a person by neural tampering,while the rest of his neural structure was left undisturbed. The person may notice the visual imaging occurring in him and realise that, because the imaging was being abnormally induced, he was not actually seeing, say, a cat jumping off a chair. But of course, if in addition all the person's thoughts were induced by neural tampering, he could not notice them or realise they were being induced - unless this last bit of cerebration was also induced.

This sort of account is of course the sort we would allow and expect of a physical structure rich enough to sustain introspective events - barring the resistance to the possibility that our mental events are physical. And we may emphasise the following points as part of the account I present:

(a)Although we make the realist distinction between events occurring in an observer and the observed events, we do not require and indeed cannot give any cogent sense to there being a "membrane or "veil" betwgen the subject (the "structure" sustaining the observing events) and the events perceived: there need only be the two events, the observing and the observed, at least the former of which occurs in the subject. If I am paying particular attention

to the perceptual events occurring in me - for example, if I open and shut my eyes and think: 'Something is happening to me when my eyes are open...' - then introspective events occur in me as well as the perceptual events. But even if it achieves anything to say that in this case my perceptual events come between my introspective events and the environmental events - which I doubt - there need only be the perceptual event and the environmental event during the process of perception.

So (b)events such as a cat jumping off a chair, the explosion of a supernova, are perceived directly (or are perceptible directly). That's to say, when, among other conditions, an event happens in the observer and another event happens, and there is a certain relation between these events, that is direct perception. The only way I can think of by which we could be in more direct contact with what we perceive is if we were what we perceived. The realist distinction need not strain the notion of direct perception.

(c)When one sees a cat jumping off a chair, we do not and must not describe this as an inferential process: say, a "subject" inferring something on the basis of his visual imagings. When we have an imaging "suggesting" a cat jumping off a chair, we do not infer from this that the cat is jumping off the chair, or infer that we are seeing a cat jumping off a chair. And if we are "wrong" - if there was no cat, and we had a visual illusion - we did not make a wrong inference, we just had a visual illusion. If we did make an inference of the sort 'There is something going on "out there in the real world"', this would be, correct or incorrect, more than we actually do when we normally have an imaging "suggesting" a cat jumping off a chair - when we see or visually illude a cat jumping off a chair.

The way the intuitive resistance to monism is to be undermined applies equally well to mental events which are not, in the loose sense so far described, designatory: perceptual events or imagings. The latter demand first attention because it is I think through them that the confused resistance to monism is forged. But if we hypothesise that such events as the occurrence of thoughts, pains, etc. may also be "suggested" or designated by imagings,

we find that we should not expect the occurrence of a pain to be like the imaging "suggesting" it.

The account which follows introduces a certain sort of formal language. Formal languages of this sort, which I call J-languages, will be applicable to perceptual events, in that the latter can be taken as syntactical entities of a J-language for which their designatory role can be defined. (But the fact that we may accordingly be able to interpret a certain event as a designating formula in a J-language should not be confused with the view we discard: that a subject interprets them.)

Much of the process of setting up the formal concepts relevant to J-languages is rather tedious; but through this process we may hopefully achieve the following:

(a)A clarification of the designatory role of imagings, which so far has only been loosely and rather misleadingly given by analogy.

(b)A rendering of the notion of Residual Privacy, which, it will be remembered, was not exponible in terms of imperceptibility or incorrigibility. But Residual Privacy will be rendered harmless to materialism.

(c)An investigation of the notions of syntactical reference and self-reference involved in this context. The possibility of syntactical reference is required for the success of the present account as a monistic theory, positing just physical events. But if the syntax/interpretation distinction is to be applied seriously here, we may expect certain consequences attending self-reference, given its pathology in formal systems theory. We will find that these consequences support and clarify the view I call subjective dualism.

4.3 A Schema for the Metatheory of J-Languages

My approach will be to extract rather abstract requirements on "observing events" without committing myself regarding their nature. Indeed, I will take the case of events as part of a more general case: in which information about something is captured by something else. The examination of this very general case will be made by introducing J-languages; sequences of events which can be interpreted as "observing events" will be taken as syntactic entities

of just one sort of J-language.

I will introduce some abstract features of J-languages, and then in 4.4 I will apply these to the case of "mental" perceptual events - again without commitment as to the exact nature of these events. But, as anticipated, we will find not only that ontological dualism cannot be justified through this application, but also that through the view of subjective dualism we can offer some explanation of the workings of the mistaken intuitive pressure for ontological dualism.[3]

4.31 A Journalistic formal language,or J-language, is conceived essentially as a formal language which records certain states of affairs. And I will call a structure which produces a record in a J-language, a J-structure.

A J-language is formal in the sense that for a given J-language, the form of the record, and the relationship between the record and what is to be recorded, are specified explicitly or implicitly so that a "machine" in the Turing sense - a structure which carries out recursive procedures - could produce the record.

There are various ways in which one could describe the formal structure of a record of the relevant sort. The way I propose to do this involves distinguishing between what will count formally as designating formulae, and what will count formally as a sequence of designating formulae. (Although there is sufficient resemblance between the normal use of the concepts of designation and truth and their use in connection with J-languages to justify the latter use, I must stress that they will be defined formally for J-languages; and no more than is given in their formal definition need be read into the use of these concepts here.)

A sequence will be a correct record if it is isomorphic (in the manner required for the record - explicitly or implicitly specified for a particular J-language) to a sequence of states of affairs to be recorded.("States of affairs", "formula", "sequence" are abstract: for example, the sequences could be spatial arrangements of objects, or events ordered in time, etc.. And the states of affairs to be recorded could be objects, events, etc..)

Each formula in a sequence will designate an element of the state of affairs to be recorded in virtue of being isomorphic to it (again in a manner particular to a given J-language). So essentially the distinction between a formula and a sequence of formulae is arbitrary - we could simply say that a correct record is isomorphic to the state of affairs recorded, whether or not we think of it as one designation or as a sequence of designations. For certain J-languages this distinction assumes greater importance; for the moment we may regard it as formally convenient to make this distinction.

The way I will handle the notion of a sequence in a J-language is this: I will distinguish between (a) open wffs of a J-language; these are taken to contain a variable, the only variable in the J-language. The variable is a positional variable, in the sense that if the variable in an open wff is replaced by a positional constant, we form a closed wff: and (b) closed wffs, to which their positional constants assign a position in a sequence of the J-language.

We could arrange that the positional constant of a closed wff also determines which sequence that closed wff is in. But instead I will make the simplifying assumption for the J-languages I will consider, that there is just one recording sequence which is generated.

A closed wff of a J-language is to be interpreted as saying that the "element" it designates occupies a position, in a sequence of states of affairs, corresponding to the position of the closed wff in its sequence. From this we get the notion of the truth of a closed wff. And the notion of derivability for a J-language is tied to whether or not the J-structure generates a certain closed wff.

No prescription need be made as to the syntactic form of the variable or the constants in open and closed wffs of a J-language. There is, however, a distinct type of J-language with which I will be primarily concerned, and which may be distinguished by an economy of syntax:- where the variable symbol of an open wff is purely notional, and where the syntactic feature giving the positional tag to a closed wff is literally the position of the closed wff in a sequence for which there is some ordering relation - for

instance, a spatial or temporal position occupied by a closed wff. My reason for concentrating on this type of J-language, with this economy of syntax, is that the "records" to which I wish to apply the notion of a J-language have this economy of syntax. One further feature of the type of case I will be primarily concerned with is that the J-structure in these cases only has the facility for producing closed wffs of the J-language. But it will be convenient nonetheless to retain the notion of an open wff for these J-languages.

The following example may help to illustrate some of the features mentioned above, and may also make it easier to follow the progress of the schema I will then give.

Suppose we wish to consider a television camera as a J-structure "using" a J-language:

(a)We may make the example more illustrative if we take the J-structure to be a camera plus screen. We may then take the closed wff to be the occurrence of a "picture" on the screen. (A photo-electronic event in the camera would do just as well, though such an event is more difficult for us to imagine and inspect.)

(b)We may note the following economy of syntax: on the intended interpretation for this J-language, a closed wff both "designates" a certain (visible) event (e.g., the scoring a goal in a football game), and"asserts" that it occurs (in a position in a sequence of visible events corresponding to the position of the closed wff in a sequence of closed wffs).

(c)Each closed wff in this J-language will form part of a sequence of such wff-events. In this case the sequence for the J-language is a sequence of events in time; and we may regard the syntactic rendering of the positional constant, determining a closed wff's position in the sequence, as being the time at which it takes place. All the closed wffs are connected in this sequence.

(d)Clearly in this simple case only closed wffs of the J-language will be generated by the J-structure. A string could not be generated which was the token of an open wff, since all wff-events generated occur at some time. But we could record the sequence of closed wffs on film, and regard

the celluloid frames in the (spatial) sequence of the film as (new) closed wffs of a new J-language. Then if we cut up the film into individual frames, each individual frame may be regarded as an open wff of this new J-language, each with a notional positional variable.

(e)Not all the events which occur in the camera or screen would be closed wffs of the J-language: the screen may explode, change in temperature, etc... The criteria of wff-hood are formally available nevertheless: in principle they are specifiable explicitly, or they may be regarded as implicitly specified by the construction of the camera and screen.

(f)It is intuitively clear how the notions of designation and truth may be made to apply in the case of the occurrence of a "picture" on the screen. But these notions will be considered in a more abstract way so that, say, designation by a photo-electronic event in the camera may be accommodated.

Before commencing the formal schema, one further point should be mentioned I will sometimes refer to the states of affairs to be recorded (or potentially recordable) also as sequences of formulae, for which we may normally have no interpretation. But, as will become apparent, we may also have an interpretation available for these.

4.32 In (1) - (5) below I will introduce some basic definitions and terminology.

(1) A J-structure generates sequences of strings; the strings generated are to be tokens of wffs in a J-language. But I will generally simplify and say that a J-structure generates sequences of wffs in a J-language. I will use the notation: the J-structure J_K generates sequences of wffs of the J-language L_K.

(2) There must accordingly be specifiable (at least in principle) criteria determining what counts as a wff of L_K. So there must be specifiable (at least in principle) an alphabet of L_K, which of course comprises the simplest syntactic elements which in allowed combinations may form wffs.

A wff of L_K is closed if it is part of a sequence in L_K; otherwise it

is open. (but a sequence may contain just one closed wff.) Accordingly there must be specifiable (at least in principle) criteria determining what counts as a sequence (of closed wffs) in L_K.

(3) I will use the following notation:

k^i is an open wff of L_K. (We may regard k^i as containing a notional positional variable.)

k_n in K' is the n-th closed wff of the sequence K'; I will sometimes omit 'in K'' as understood, since we need only concern ourselves with one sequence. (We may regard k_n as containing a positional constant determining that it is the n-th closed wff (in K').)

k_n^i in K' is the closed wff formed when k^i, with a notional positional variable, is given a constant ("replacing" the notional variable) determining that it is in the n-th position in K'.

(4) The J-language L_K is the set of all its open and closed wffs. K, the syntactical domain of L_K, is the set containing all and only the open wffs of L_K.

K_α is the set of the alphabet "letters" of L_K. The individual alphabet "letters" of L_K will be specified as α_1^K, α_2^K, etc..

(5) W, the domain of interpretation of L_K, is the set of objects sequences of which L_K is intended to record. I will say then that L_K is a J-language for W. And I will use the following notation:

w^j is a member of W.

w_n in W' occupies the n-th position in the sequence W' of members of W.

w_n^j in W' is the w^j in the n-th position of the sequence W'.

W_α is the set of "components" of members of W: I am assuming that each member of W can be rendered as a combination of primitive components.

(6) I will now describe designation for a J-language, in terms of isomorphism. The notion of isomorphism may be intuitively grasped in terms of "common structure". But I will use a set-theoretical definition of isomorphism (6a),

in terms of which I will define (in 6b) designation for a J-language: J-designation. Intuitively we may regard the relationship between a wff of a J-language and its (J-) designatum as involving common structural features. In (6c) I indicate formally in terms of relations how the structural features of the domain of interpretation are captured by the J-language.

In (6d) I give an example which illustrates most of the terminology introduced so far.

(6a) Isomorphism may be defined set-theoretically as follows:[4]

$(X;R_0, R_1, \ldots, R_p)$ is a "system" if X is a set, and $R_0, R_1, \ldots, R_p$ are given relations on X.

Let $(X;R_0, R_1, \ldots, R_p)$ and $(Y;S_0, S_1, \ldots, S_p)$ be two systems, the relations R_t and S_t being of the same kind for each t : namely both (n_t+1)-ary.

Then a mapping $f:X\to Y$ is said to be an isomorphic mapping of the first system into the second if it is a bijection from X to Y - if it is a mapping of X into Y (a function from X to Y) putting the members of X and Y in one-to-one correspondence - with the property that, for $t = 0,\ldots,p$,

$x_0, \ldots, x_{n_t} \in X \supset (R_t(x_0, \ldots, x_{n_t}) \text{ iff } S_t(f(x_0), \ldots, f(x_{n_t})))$.

When $(X;R_0, \ldots, R_p)$ is mapped isomorphically by f into $(Y;S_0, \ldots, S_p)$, we write $(X;R_0, \ldots, R_p) \overset{f}{\simeq} (Y;S_0, \ldots, S_p)$.

I will abbreviate by having $\underline{R}_X$ be the set whose members are $R_0, \ldots, R_p$, (the given relations on X); and $\underline{R}_Y$ be the set whose members are $S_0, \ldots, S_p$, (the given relations on Y). Then we specify the two systems as $(X;\underline{R}_X)$ and $(Y;\underline{R}_Y)$. And if there is at least one isomorphic mapping of $(X;\underline{R}_X)$ into $(Y;\underline{R}_Y)$, we write: $(X;\underline{R}_X) \simeq (Y;\underline{R}_Y)$.

(6b) We now get our rule of designation in L_K as follows:

There must be a function $f:K_\alpha \to W_\alpha$, a bijection $j:K \to W$, some set $\underline{R}_{K_\alpha}$ of relations on K_α, and some set $\underline{R}_{W_\alpha}$ of relations on W_α, such that for each i,

$k^i \in K \supset ((k^i;\underline{R}_{K_\alpha}) \overset{f}{\simeq} (j(k^i);\underline{R}_{W_\alpha}))$.

For a given J-language L_K, f, j, $\underline{R}_{K_\alpha}$ and $\underline{R}_{W_\alpha}$ must be in principle specifiable.

Then k^i designates $j(k^i)$ in L_K.

I will use the following notation:

The relation $\overset{K\to W}{J}$ is such that $\overset{K\to W}{J}(k^i, w^j)$ iff k^i designates w^j in L_K, i.e., $\overset{K\to W}{J}(k^i, w^j)$ iff $j(k^i) = w^j$. And we may also say that $k^i = k(w^j)$ iff $j(k^i) = w^j$.

(6c) It may be intuitively grasped that since <u>for each</u> i, $(k^i; \underline{R}_K) \overset{f}{\simeq} (j(k^i); \underline{R}_W)$, (roughly: k^i is "isomorphic to" $j(k^i)$), then for some $\underline{R}_K$ and $\underline{R}_W$ the bijection $j: K \to W$ is an isomorphic mapping of $(K; \underline{R}_K)$ into $(W; \underline{R}_W)$: $(K; \underline{R}_K) \overset{j}{\simeq} (W; \underline{R}_W)$. I will not give a proof of this. (Here and elsewhere where there is an intuitive account whose rigorous treatment requires rather cumbersome detail, I will sacrifice the rigour in order to cover the ground required in the present context.) Instead I give an example. Suppose we have two packs of playing cards. If we know that for each card in the first pack there is one and just one isomorphic card in the second pack, we can see that there must be an isomorphism between the two packs. The analogy is between the two packs and K and W; and between a card from each pack and k^i and w^j. (Thus in order to give precision to the notion of isomorphism between the cards we would analogously have to take the cards as ordered sets.)

If we return to the definition of isomorphism we can get the notion of "corresponding relations" in isomorphic systems.

When $(X; R_o, \ldots, R_p) \overset{f}{\simeq} (Y; S_o, \ldots, S_p)$, as defined in (6a), I will say that S_t is the corresponding relation to R_t, and that $S_t = f\text{-}R_t$.

So in the case where $(K; \underline{R}_K) \overset{j}{\simeq} (W; \underline{R}_W)$, I will say that $\overset{\text{if}}{\wedge} R_t \in \underline{R}_K$, $j\text{-}R_t$ is its corresponding relation. And if, for example, R_t is a binary relation, then $R_t \in \underline{R}_K \supset (R_t(k^i, k^j)$ iff $j\text{-}R_t(j(k^i), j(k^j))$.

(6d) <u>Example</u>: (I improperly use the general terms 'L_K','K','W', etc., to apply in the case of the particular J-language in the example. I do this in order to exemplify more vividly the notions and notation used so far. But in future if I discuss a particular J-language, unless in a clarifying example, I will use, say, 'L_A','A','A_α','a^i', etc.. Also, in the example below, quotation marks round the symbols '0','1','a','b' are omitted throughout for visual clarity.)

$K_\alpha = \{0,1\}$. So the alphabet of L_K is: 0,1 . And $\alpha^K_1 = 0, \alpha^K_2 = 1$. I will use 'λ' and 'μ' as variables instead of the numerical subscripts specifying the alphabet "letter".

$W_\alpha = \{a,b\}$.

$K = \{01,10,00,11\}$. And $k^1 = 01, k^2 = 10, k^3 = 00, k^4 = 11$.

$W = \{\overset{a}{b}, \overset{b}{a}, \overset{a}{a}, \overset{b}{b}\}$. And $w^1 = \overset{a}{b}$, etc..

Sequence in L_K: open wffs become closed if written in a vertical list, the n-th closed wff in a sequence being the n-th closed wff from the top of the list.

$f: K_\alpha \rightarrow W_\alpha$ is such that $f(0) = a, f(1) = b$.

$\underline{R}_{K_\alpha} = \{H\}$ where $H(x,y)$ iff x is written horizontally to the left of y.

$\underline{R}_{W_\alpha} = \{V\}$ where $V(x,y)$ iff x is written vertically above y.

$j: k \rightarrow W$ is such that, for each i (for $i = 1,..,4$),

if $k = \langle \alpha^K_\lambda, \alpha^K_\mu \rangle$, then $j(k^i) = \langle f(\alpha^K_\lambda), f(\alpha^K_\mu) \rangle$, and $H(\alpha^K_\lambda, \alpha^K_\mu)$ iff $V(f(\alpha^K_\lambda), f(\alpha^K_\mu))$.

So (for each i) $k^i \in K \supset ((k^i; \underline{R}_{K_\alpha}) \overset{f}{\simeq} (j(k^i); \underline{R}_{W_\alpha}))$.

For example, k^1 designates w^1 in L_K.

Then suppose W' were: $\overset{a}{a}\ \overset{b}{a}\ \overset{a}{a}\ \overset{b}{b}$ (Strictly we should specify what is to count as a sequence of members of W to be recorded).

J_K would then generate or write as K' :

00 (k^3_1)
10 (k^2_2)
00 (k^3_3)
11 (k^4_4)

(This is providing J_K was veridical - I discuss this below. I also discuss the notion of truth-in-a-J-language below. The following closed wffs would turn out to be true in L_K : k^3_1 , k^2_2 , k^3_3 , k^4_4 .)

(7) I will now define for J-languages the notions of derivability and truth, and the notion of a veridical J-structure. Intuitively the notion of truth is justified in application to J-languages, in that we may interpret a closed wff of a J-language as saying something about the domain of interpretation of the J-language.

(7a) The notion of the derivability of a closed wff is simply connected to the notion of whether it is generated by the J-structure. I should point out here that we suppose that there is some way by which W' is generated - indeed, we may suppose that there may be a structure J_W which embodies some set of rules and thereby generates W'. But how W' is generated need not concern us.

Now, I will say that $D(k^i_n)$ iff k^i_n is generated by J_K as the n-th closed wff of K'.

And I will say that k^i_n is <u>derived</u> by J_K in L_K iff $D(k^i_n)$ and n is finite. We are of course asuming that K' is the only sequence generated by J_K. In general we will continue to make this assumption, but for the definition of <u>derivability</u> we must consider <u>possible</u> sequences which could be generated by J_K and say:

k^i_n is <u>derivable</u> by J_K in L_K iff k^i_n <u>could</u> be generated by J_K as the n-th closed wff of <u>some</u> sequence K' (and n is finite).

(7b) k^i_n is <u>true</u> in L_K iff $w^j_n = j(k^i_n)$. And k^i_n is <u>false</u> in L_K iff $w^j_n \neq j(k^i_n)$. Note that in this definition of truth for a J-language the truth of a closed wff is independent of whether or not it is derivable.

(7c) A J-structure J_K is <u>veridical</u> in L_K if, when J_K is operating, for each n and for each i, $D(k^i_n)$ iff k^i_n is true in L_K.

(7d) <u>Comment</u>: If J_K is veridical (in L_K) we may regard it as designed to "capture" certain facts "mechanically". And, if J_K is veridical, then without being able to say what W' will be like, we can guarantee that J_K will accurately record W'. So although the wffs of L_K which are derived by J_K cannot be specified independently of the sequence W' from the domain of interpretation W, a certain "formal apparatus" is nonetheless embodied in the J-structure independent of the particular sequence W' to be recorded.

I will be concerned mainly with veridical J-structures.

(8) The concepts I now introduce, centering around a notion of "accessibility", concern what may be regarded as the "amount of information" about a designatum which is captured by its designating wff.

(8a) W is <u>accessible</u> to L_K <u>relative to</u> $\underline{R}_{W_\alpha}$ if, for some $\underline{R}_{K_\alpha}$, $f:K_\alpha \rightarrow W_\alpha$, $j:K \rightarrow W$, and for each i, $k^i \in K \supset ((k^i;\underline{R}_{K_\alpha}) \overset{f}{\simeq} (j(k^i);\underline{R}_{W_\alpha}))$.

W is accessible to L_K, <u>through</u> J_K, relative to $\underline{R}_{W_\alpha}$, if W is accessible to L_K relative to $\underline{R}_{W_\alpha}$, and for any sequence W', for each n, $D(k(w_n))$; (i.e., $D(k_n)$ and $J^{K \rightarrow W}(k_n,w_n)$). In this case we may also say that W is accessible to J_K <u>in</u> L_K.

So, if J_K is veridical in L_K (for W), then W is accessible to L_K through J_K relative to some $\underline{R}_{W_\alpha}$.

If W is accessible to L_K relative to $\underline{R}_{W_\alpha}$, we may also say that W is accessible to K relative to $\underline{R}_{W_\alpha}$. And if for each i, $k^i \in K \supset ((k^i;\underline{R}_{K_\alpha}) \overset{f}{\simeq} (j(k^i);\underline{R}_{W_\alpha}))$, we may also say that W is accessible to L_K (or K) <u>with</u> $\underline{R}_{K_\alpha}$, relative to $\underline{R}_{W_\alpha}$.

(8b) Let $\underline{R}^T_{W_\alpha}$ be the set of all relations on W_α and $\underline{R}^T_{K_\alpha}$ be the set of all relations on K_α.

Then W is <u>totally</u> accessible to L_K iff W is accessible to L_K relative to $\underline{R}^T_{W_\alpha}$. If W is accessible to L_K only relative to a (proper) subset of $\underline{R}^T_{W_\alpha}$, then W is <u>partially</u> accessible to L_K.

(8c) W is <u>univocally</u> accessible to L_K with $\underline{R}_{K_\alpha}$ relative to $\underline{R}_{W_\alpha}$ iff W is accessible to L_K (with $\underline{R}_{K_\alpha}$) relative to $\underline{R}_{W_\alpha}$, and there is only one function $f:K_\alpha \rightarrow W_\alpha$ such that for each i, $k^i \in K \supset ((k^i;\underline{R}_{K_\alpha}) \overset{f}{\simeq} (j(k^i);\underline{R}_{W_\alpha}))$; (in which case there is only one function $j:K \rightarrow W$, since j is determined by f). Otherwise, if there is more than one such function f, W is only <u>equivocally</u> accessible to L_K relative to $\underline{R}_{W_\alpha}$.

(8d) <u>Comment</u>: It should be evident that in the case of the simple J-language of example (6d), W was equivocally partially accessible to L_K (or to J_K in L_K).

Where a domain W is totally accessible to a J-language, in effect <u>every</u> feature of a member of W is <u>mapped</u> by a feature of its designating wff. And if W is only partially accessible to J_K, then there are features of w^i which $k(w^i)$ does not map.

But even if W is totally accessible to J_K, there may still be a sense

in which not <u>every</u> feature of w^i is <u>captured</u> by $k(w^i)$, even though every feature of the former is <u>mapped</u> by the latter: if there is a w^j which is "totally isomorphic" to w^i (in which case, as we shall see later, W is not univocally accessible to J_K), then if $k^i = k(w^i)$, k^i would also map every feature of w^j. So k^i would not then capture the "feature" of w^i that it is w^i and not w^j. So we can see that in this case the information contained in a wff about its designatum is in a sense "incomplete". These considerations motivate the definition of a concept I will call <u>J-privacy</u>.[5]

(9) W is <u>J-private</u> from J_K in L_K iff W is not univocally accessible to J_K in L_K.

Consider L_K such that $K' = W'$. Then when $j(k^i) = k^i$, for each i,

$k^i \in K \supset ((k^i;\underline{R}^T_{K_\alpha}) \overset{F}{\simeq} (j(k^i);\underline{R}^T_{K_\alpha}))$ ...(i)

So K is totally accessible to J_K in L_K.

Now, even if there is more than one $j:K\rightarrow K$ such that (i) obtains, there can only be one $j:K\rightarrow K$ such that (i) obtains <u>and</u> $D(k_n) \supset (j(k_n) = w_n)$; namely where $j(k_n) = k_n$. So K is univocally totally accessible to J_K in L_K.

Thus we have <u>Theorem 1: If L_K is such that $K' = W'$, K is not J-private from J_K in L_K.</u>

(10) Suppose that W is totally accessible to L_K. Then any two objects "made from" W_α are designated by two different wffs of L_K. Some further concepts which will be required involve the question of whether any two different formulae "made from" K_α designate two different objects from W: part of the alphabet of L_K, or part of the set $\underline{R}^T_{K_\alpha}$, may be redundant as far as mapping W is concerned. To formalise this question I will introduce the concepts of <u>efficiency</u>, <u>continuity</u>, and <u>graphicity</u> for J-languages.

(10a) Suppose W is accessible to K relative to $\underline{R}_{W_\alpha}$. Then if K is totally accessible to W with $\underline{R}_{W_\alpha}$, L_K is <u>efficient</u>; (and if K is only partially accessible to W with $\underline{R}_{W_\alpha}$, L_K is <u>inefficient</u>).

<u>Comment</u>: If K is totally accessible to W with $\underline{R}_{W_\alpha}$, then in effect every feature of a wff of L_K is put to use when designating, since every feature of k^i is mapped by a feature of $j(k^i)$ within the set of features of $j(k^i)$ which k^i is intended to capture. This is a sufficient condition for any two

formulae "made from" K_α to designate two objects "made from" W_α.

(10b) Let $\underline{R}_{K_\alpha}$ be the set of relations $\{R_o,\ldots,R_p\}$.

And let $\Delta\underline{R}_t$ be the set of all things which can partake in the relation R_t; (e.g., if R_t is binary, ΔR_t is the union of the domain and range of R_t).

And let $\Delta\underline{R}_{K_\alpha}$ be the set of all things which belong to at least one of the sets in the family of sets $\Delta R_o,\ldots,\Delta R_p$; (i.e. $\Delta\underline{R}_{K_\alpha}$ is the union of that family of sets).

Now, if for any wff k^i of L_K, and for any formula F constructed out of members of $\Delta\underline{R}_{K_\alpha}$, F is <u>another wff</u> of L_K provided it is different from k^i relative to $\underline{R}_{K_\alpha}$, then L_K is <u>continuous relative to</u> $\underline{R}_{K_\alpha}$.

(10c) If L_K is continuous relative to $\underline{R}^T_{K_\alpha}$, then L_K is <u>graphic</u>.

(10d) <u>Comment</u>: If L_K is efficient, then even though k^i is such that every feature of it is used to map some feature of $j(k^i)$, a formula slightly different from k^i may not designate a correspondingly different object from W: it may not be a wff of L_K. But if L_K is continuous relative to $\underline{R}^T_{K_\alpha}$, this guarantees that however slightly a formula (an object within the alphabetic confine for the J-language) differs from a wff k^i, it is still a wff of the J-language; and this slightly different wff will designate a w^j correspondingly different from $j(k^i)$. So without giving a rigorous proof we can see that if L_K is graphic it is also efficient.

The concepts of efficiency and continuity essentially reflect the economy with which L_K uses its syntactic domain. A graphic J-language uses its syntactic domain as economically as possible, and in this sense is a limiting case. I will discuss this again later.

(11) I will call $m:W\rightarrow W$ an <u>inverting function</u> for W if for each i,

$$w^i\epsilon W \supset ((w^i;\underline{R}^T_{W_\alpha}) \simeq (m(w^i);\underline{R}^T_{W_\alpha})). \qquad \ldots(i)$$

And evidently, if (i) obtains, then for each i,

$$w^i\epsilon W \supset ((w^i;\underline{R}_{W_\alpha}) \simeq (m(w^i);\underline{R}_{W_\alpha})). \qquad \ldots(ii)$$

(11a) <u>Lemma 1</u>: If there is an inverting function for W, then W is not univocally accessible to J_K in L_K.

Proof: Suppose W is accessible to J_K in L_K relative to $\underline{R}_{W_\alpha}$; then for each i,

$k^i \in K \supset ((k^i ; \underline{R}_{K_\alpha}) \simeq (j(k^i); \underline{R}_{W_\alpha}))$. ...(iii)

(And for each n, $D(k_n)$ such that $w = j(k_n)$).

Now, if we substitute $j(k^i)$ for w^i in (ii), then it follows from (iii) by the transitivity of isomorphism (if $(X;\underline{R}_X) \simeq (Y;\underline{R}_Y)$ and $(Y;\underline{R}_Y) \simeq (Z;\underline{R}_Z)$, then $(X;\underline{R}_X) \simeq (Z;\underline{R}_Z)$) that there is another function $j';K\to W$ such that for each i,

$k^i \in K \supset ((k^i ; \underline{R}_{K_\alpha}) \simeq (j'(k^i); \underline{R}_{W_\alpha}))$, ...(iv)

namely where $j'(k^i) = m(j(k^i))$. And J_K could equally well operate this j-function in L_K. That's to say, given exactly the same L_K (the same set of open and closed wffs), J_K could instead of operating j operate j' such that (iv) obtained and, for each n, $D(k_n)$ such that $w = j'(k_n)$. And we can extend this result for any $K(\neq W)$ and obtain

Theorem 2: If there is an inverting function for W, then W is not univocally accessible to any J_K (where $K \neq W$).

And if there is any inverting function for W we may say that W is J-private to J_W.

(11b) Theorem 3: If L_K is a graphic J-language for W, (and $W \neq K$), then if there is an inverting function for K, W is J-private from J_K in L_K.

Proof: Since L_K is graphic, for each i,

$k^i \in K \supset ((k^i ; \underline{R}^{\tau}_{K_\alpha}) \simeq (j(k^i); \underline{R}_{W_\alpha}))$. ...(i)

The inverting function for K, $m:K\to K$, is such that for each i,

$k^i \in K \supset ((k^i ; \underline{R}^{\tau}_{K_\alpha}) \simeq (m(k^i); \underline{R}^{\tau}_{K_\alpha}))$. ...(ii)

Since L_K is graphic, if k^i is a wff of L_K so is $m(k^i)$. So we can substitute $m(k^i)$ for k^i in (i), and then it follows from (ii) by the transitivity of isomorphism that there is another function $j';K\to W$ such that for each i,

$k^i \in K \supset ((k^i ; \underline{R}^{\tau}_{K_\alpha}) \simeq (j'(k^i); \underline{R}_{W_\alpha}))$,

namely where $j'(k^i) = j(m(k^i))$. And J_K could operate either function j or j' in L_K.

(11c) Comment: The salient point about Theorem 3 is this: to determine whether or not W is totally accessible to J_K, we have to determine the j-function - in effect, we have to compare W and K. But, from Theorem 3, W can

be determined to be only equivocally accessible to J_K without determining the j-function - without comparing W and K-merely by ascertaining whether there is an inverting function for K. And if there is, and W is J-private from J_K, then unless we can determine the j-function which J_K is operating (- in effect: unless we can compare K' an W' -) then we cannot uniquely determine the designatum of k^i.

(12) In an ordinary language, or an ordinary formal language, the relation between the syntax of the language and its interpretation is essentially a conventional one. The syntactical form of the alphabet of a conventional language L is arbitrary: it doesn't matter what properties two letters of the alphabet have, so long as they are different. In this sense the relation between the syntactic atoms of L and the intended interpretation of their combinations is essentially conventional. And although there must be some logical restrictions imposed by the choice of interpretation, the syntactic atoms of L can be described logically independently of their role in L - described, that is, qua objects rather than qua symbols.

The arbitrariness of the symbols in conventional languages reflects itself also in what may be described as a discontinuity in the alphabet letters of such a language. This may also be present in J-languages. Discontinuity in a J-language which is continuous relative to $\underline{R}_{K_\alpha}$ is reflected in the "discrepancy" between the sets $\underline{R}_{K_\alpha}$ and $\underline{R}^{\tau}_{K_\alpha}$: the relations on the syntax which are not used in capturing features of the domain of interpretation are "redundant". In the case of a graphic J-language, this discrepancy is erased, and as a result the open wffs of such a J-language cannot be broken down into syntactic atoms whose properties qua objects are arbitrary, and are not restricted by the interpretation of the language. This is the striking respect in which designation in a graphic J-language presents a limiting case.

Thus if k^i is a wff of a graphic J-language, there are no features of k^i which are superfluous for its designatory role. But it should not be supposed that there cannot therefore be superfluous and non-designatory features

of a sequence in a graphic J-language - there may be; graphicity as I have defined it is only applicable in respect of "redundancy" to open wffs of a J-language.

There is another sort of J-language which presents a limiting case. If L_K is a J-language whose closed wffs are given their positional tag by their ordering in time - its closed wffs are events - then I will call L_K a temporal J-language. In this case there can only be notional open wffs in such a J-language, in the sense that a string which is the token of an open wff can only be notional. Nevertheless we may usefully retain the concept of an open wff (and the set K of these) for a temporal L_K : the open wffs of a temporal J-language may be treated rather as are the logical states of a physically realised Turing machine. Thus the passage of a physically realised Turing machine through a finite number of logical states may nevertheless be infinitely divisible. Similarly, a temporal sequence in a J-language may be infinitely divisible in time, though comprising only a finite number of closed wffs - each of which may be regarded as an open wff given a positional tag.

(This is rather makeshift and sketchy - but there are issues involved which though too deep for further examination in this context nevertheless should be aired.)

(13) We may consider the case where J_K is extended so that it uses a conventional language. For this we can augment J_K in a fairly obvious way and introduce:-

(a) A scanner for J_K . By means of the scanner J_K can scan different "regions" of a sequence W' from the domain W, or different sequences from that domain.

We may distinguish between the case where W', or a "region" of W', is not being scanned; and the case where a sequence U', not being from the domain W, is not captured by L_K . In the first case we may say that W' (or the "region" of it) is not recorded by J_K ; and in the second case that U' is not, at least directly, recordable by J_K . I will also describe, say, J_K recording w_n , as J_K perceiving w_n . But I do not intend this special usage to suggest that it conforms in general with the ordinary concept of perception; (though in some

cases, for some J-languages and J-structures, it will be intended to conform with the ordinary concept of perception).

(b) A manipulator for J_K , by means of which J_K can alter the sequence W'. It is then possible for J_K to "write" an alphabet superimposed on K': an alphabet of a language L might comprise symbols which qua objects are tokens of wffs of L_K , but which in addition are tokens of the conventional language L. And an expression of L may then be construed by J_K manipulating W' so that terms of L are generated in the sequence K'.

Example: (I omit quotation marks round '0' and '1'.)

$K = \{0,1\}$. Sequence of L_K : a horizontal line of wffs sepated by commas. We need not concern ourselves with the nature of W', except that $w_n = w_{n-1}$ unless J_K intervenes with its manipulator.

Suppose K' is : 0,0,1,1,0

Then J_K has altered W' by means of its manipulator at the third and fifth positions of W'. Now, in addition to the J-linguistic interpretation, we may interpret K' as designating in a language L the number 2.

(14) I will now introduce the notions of a secondary interpretation for a J-language, and a secondary J-language. But as we shall see we should not confuse these notions with the possible availability, mentioned just now in (13), of interpretations which are not J-linguistic.

Suppose a domain U is not accessible to a J-structure J_K . We may say then, as suggested in (13a), that a sequence U' from U is not perceptible to J_K.

But now consider the following case. L_K is a J-language for W; but there is a subset X of W, which is the set of open wffs of a J-language L_X for U, there being a J-structure J_X . Now, L_X is such that, by definition, for each j , ${}^{X\to U}J(x(u^j), u^j)$, where $u^j \in U$.

And for a subset K_1 of K, we have a relation ${}^{K_1\to X}J$, such that for each j , ${}^{K_1\to X}J(k(x(u^j)), x(u^j))$.

And since ${}^{K\to W}J$ is the function j: K→W, ${}^{K_1\to X}J$ is a restriction of the function j; (${}^{K_1\to X}J$ is the function $j|K_1$, where $j|K_1 : K_1 \to X$ is a function formed by restricting the domain of j to K_1).

So we have a resultant relation $J^{K_1 \to U}$ such that for each j , $J^{K_1 \to U}(k(x(u^j)),u^j)$.

This forms a <u>secondary interpretation</u> for certain wffs of L_K: for each i , $k^i \in K \supset (k^i$ designates $x(u^j)$ in L_K iff k^i designates u^j in $L_{K_1})$.

In other words, we have two rules of designation for k^i (if $k^i \in K_1$), one given by the function $J^{K \to W}$, the other by the function $J^{K_1 \to U}$. And if $K_1 \neq K$ (i.e., if K_1 is a proper subset of K), we have the <u>secondary</u> J-language L_{K_1} .

If J_X is part of the causal chain from U' to K', then we may say that J_K may <u>indirectly perceive</u> U'. And we may call X an <u>amplifying domain</u> for U, X' an <u>amplifying sequence</u>, and J_X an <u>amplifying J-structure</u> for U.[6]

Thus any distinct function (giving a distinct rule of designation for L_K) which J_K may be said to be operating as it generates K' gives a distinct interpretation for L_K . We can see then that the following secondary interpretation is always available for a J-language L_K for any W (with any W as the primary - roughly: the intended - interpretation): in (9) we came across the function $j:K \to K$ such that $j(k^i) = k^i$. I will call the function $n:K \to K$, where $n(k^i) = k^i$, the <u>null function</u> for K. We can thus form the <u>null interoretation</u> for any L_K , given by the rule of designation: for each i ,

$$k^i \in K \supset ((k^i ; \underline{R}^T_{K_\alpha}) \simeq (n(k^i) ; \underline{R}^T_{K_\alpha}));$$

and of course, whatever the primary j-function j for L_K ,

$(D(k_h)$ such that $j(k_n) = w_n) \supset (D(k_n)$ such that $n(k_n) = k_n)$. The secondary interpretation is then the null interpretation, whose domain is K.

(15) We may now reconsider the notion of truth for J-languages, and introduce the notions of (semantic) completeness and (semantic) consistency for J-languages.

(15a) I will call truth for J-languages as defined in (7b) <u>truth'</u>:

k_n is <u>true'</u> in L_K for W iff $w_n = j(k_n)$.

(I insert 'for W', which was ommitted in (7b), in order to accommodate the possibility of different interpretations for L_K .)

This conforms with the idea that a closed wff is "true" or "false" independently of whether it is generated, indeed, of whether J_K is operating. But there is another formulation of truth for a J-language which conforms to this intuitive basis: where a closed wff k_n is taken as "true" provided that if it were generated by J_K , $j(k_n) = w_n$. We thus have <u>truth</u>^{2}:

k_n is <u>true</u>^{2} in L_K for W iff ($D(k_n) \supset (j(k_n) = w_n)$);

(i.e., iff ($D(k_n) \supset (k_n$ is true$'$ in L_K for W))).

Clearly, if k_n is true$'$ in L_K for W, then it is true^2 in L_K for W. And in most cases we should expect equivalence between truth$'$ and truth^2 . However, there is a case where truth$'$ and truth^2 diverge. I will only briefly remark on this, as a fuller discussion belongs to a region of the philosophical logic of J-languages which lies outside my present scope.

Suppose we consider the null interpretation for a J-language L_K : i.e., we consider L_K as a J-language for K. It may be argued that in this limiting case the concept of truth ceases to apply. Nevertheless, if we do apply it we find that: suppose K' comprises just one closed wff, k_i^i ; then k_i^i is true$'$ in L_K for K (in L_K on the null interpretation); and every other closed wff is false$'$. Thus, for example, k_i^j is false$'$ in L_K on the null interpretation - yet, if it had been generated instead of k_i^i , it would have been true$'$. So k_i^j , and indeed every closed wff of L_K on the null interpretation, is true^2 . (The fact that $n: K \to K$ cannot furnish a criterion of generation for J_K does not vitiate this consideration:- certainly J_K cannot generate K' according to just the rule: $D(k_n)$ iff $n(k_n) = k_n$; nevertheless the null interpretation is available where there is a j-function such that L_K has a domain of interpretation W such that W' $\neq$ K'.)

(15b) J_K is <u>complete</u>$'$ in L_K for W if, when J_K is operating, for each n, w_n is generated as the n-th member of $W' \supset (D(k_n)$ such that $j(k_n) = w_n)$.

J_K is <u>complete</u>^{2} in L_K for W if, when J_K is operating, for each n, w_n is generated as the n-th member of $W' \supset ((D(k_n) \supset (j(k_n) = w_n)) \supset D(k_n))$.

The null interpretation for L_K again divides completeness$'$ and completeness^2: if K' comprises just k_i^i , then J_K is complete$'$ on the null interpretation (where W' = K'); but J_K would not be complete^2 (would be incomplete^2) on the

null interpretation for L_K.

(15c) J_K is <u>consistent</u> in L_K for W if, for each n,

$D(k_n) \supset (j(k_n) = w_n)$; (i.e., if $D(k_n) \supset (k_n$ is true' in L_K for W).

Clearly a veridical J-structure (as defined in (7c)) is complete' and consistent in L_K for W.

(15d) J is <u>j-consistent</u> for the domains of interpretation $W_o, \ldots, W_\rho$, if for each n, for each t , and for each j ,

$D(k_n)$ and $j(k_n) \in W_t \supset k_n$ is true' in L_K for W_t .

Thus, for instance, if J_K is consistent in L_K for W ($\neq$ K), then J_K is j-consistent for the domains W and K (the latter being the null domain of interpretation).

(16) Suppose J_K is a veridical J-structure in L_K for W. And suppose we wish to investigate by means of J_K whether $w_m = w_n$. We can do this by determining whether $k(w_m) = k(w_n)$, since $w_m = w_n$ iff $k(w_m) = k(w_n)$. In other words the fact that $w_m = w_n$ reflects itself in the fact that $k(w_m) = k(w_n)$.

But suppose that K <u>is a subset of</u> W. Then we have the possibility that $w_m = k_n^i$. This possibility reflects itself in a rather different manner: for $w_m = k_n^i$ iff $j(k(w_m)) = k_n^i$. Thus the "identity hypothesis" that $w_m = w_n$, and the "identity hypothesis" that $w_m = k_n^i$ reflect themselves rather differently as regards determination by J_K. In contrast to the identity '$w_m = w_n$', which is reflected by the identity relation obtaining between $k(w_m)$ and $k(w_n)$, the "unusual" identity '$w_m = k_n^i$' is reflected not by the identity relation, but by the relation $\overset{K \to W}{J}$ obtaining between $k(w_m)$ and k_n^i .

When we apply J-linguistic concepts to "mental" perceptual events, this contrast will form part of the germ of the thesis of subjective dualism: for the "unusual" identity hypothesis, whose truth requires the obtaining of a certain relation between two <u>different</u> wffs, will roughly transpose to the identity hypothesis that a perceived event is a "mental" perceptual event. But where $w_m = k_n^i$, the relation between $k(k^i)$ and k^i is not identity provided that $k(k^i)$ does not J-designate itself. We will also see that, except on the null interpretation, a self-designating wff cannot be generated.

4.4 The J-Linguistic Treatment for Mental Perceptual Events.

In 4.41 and 4.42 I will consider part of the human neural structure as a J-structure J_A ,in which visual perceptual events are taken to occur. (Again I take visual perceptual events for simplicity - they suffice for the points I wish to make). J_A will be assumed to be part of a much more complex structure J_A^+ , in which not merely perceptual events occur; but intially we need not take this into account.

4.41 (1) The domain M is a set of physical events, and M' is a temporal sequence of members of M.

(2) V is a (proper) subset of M comprising just visible events; and V' is a (temporal) sequence of members of V.

(3) A' is a sequence of events, members of the domain A, which is generated by the J-structure J_A , such that the events in A' are closed wffs of a temporal J-language L_A for the domain V.

(4) J_A is assumed to be "normally" a veridical J-structure in L_A for V. And L_A will be assumed to be a graphic J-language. (The latter is rather a makeshift assumption. The reason why this assumption is made is that, at least introspectively, a J-language for which visual perceptual events are construed as wffs would seem to be efficient and continuous. Some questions relating to this assumption will be mentioned in Part 5.)

(5) A' is assumed to be generated in the same manner as M': in virtue of causal interactions. And A is assumed to be a (proper) subset of M. (This is essentially a monistic hypothesis. It should be noted though that dualism would not simply entail taking M and A as exclusive of each other - since the generation of A' would then require comment.)

(6) If a_n is the n-th wff-event of A', and v_n is the n-th member of V' (which is scanned by J_A) then since J_A is veridical, for each n,

$J^{A \to V}(a_n, v_n)$; and a_n J-designates v_n in L_A .

J_A is assumed to operate the j-function, and so maintain an isomorphism, in virtue of causal interactions between V' and J_A .

(7) In the present view we hypothesise that this account at least roughly describes the case where A is the set of possible visual imagings that may occur in a subject, and A' a sequence of visual imagings. And we also hypothesise that A' is also describable as a sequence of neural events. For the moment we shall continue exhibiting some J-linguistic features involved. But we shall soon see that a J-linguistic rendering of the monism/dualism issue (applied to the case of visual imagings/neural events) reveals ontological monism as unobjectionable, and ontological dualism as unjustified.

(8) We can easily suggest an inverting function for A,m:A→A : - where if a^i is a certain visual image, $m(a^i)$ is the lateral inversion of a^i . (My use of 'image' rather than 'imaging' is prompted by the fact that both an open wff a^i and a visual image (in a sense abstracted from 'visual imaging') are notional entities - L_A being a temporal J-language (cf. 4.32,(12)). In this case they are roughly one and the same entity - in other words, we may usefully treat the relation between a visual imaging and the appropriate visual image as the relation between a closed wff and an open wff of the relevant temporal J-language.) By Theorem 3 (4.32,(11b)) this is sufficient to show that V is J-private from J_A in L_A .

But although V is J-private from J_A in L_A , we may of course say that J_A perceives V' directly. Furthermore, J_A may indirectly perceive an event from a domain other than V if there is a suitable amplifying sequence for that domain. (My use of 'perceive' here is intended to fuse its use as defined J-linguistically with its ordinary use.)

(9) Suppose J_B is a J-structure similar to J_A , and L_B is also a temporal J-language for V. We have then that for each n,

$J^{B\rightarrow V}(b_n, v_n)$, where b_n is the n-th wff-event of B'.

Now, we will assume that A is not a subset of V. So J_B cannot directly perceive A'. But suppose there is an amplifying J-structure J_X for the domain

A, such that the amplifying domain X is a subset of V. So L_X is a J-language for A. Then , for each n, (assuming of course that J_X is veridical in L_X), $J^{X\to A}(x_n, a_n)$, (where x_n is the n-th wff of X'); and (assuming J_B is scanning X'): $J^{B_1\to X}(b(x_n), x_n)$, where B_1 is a subset of B, and $J^{B_1\to X}$ is a restriction of $J^{B\to V}$. So for each n we have the resultant:

$J^{B_1\to A}(b(x_n), a_n)$.

This gives L_{B_1} as a secondary J-language for A; and A' is indirectly perceptible to J_B.

And just as we form L_{B_1}, we can form L_{A_1}. Thus, assuming J_A is scanning X', then for each n, $J^{X\to A}(x_n, a_n)$, $J^{A_1\to X}(a(x_n), x_n)$ and $J^{A_1\to A}(a(x_n), a_n)$; A_1 being a subset of A. (Of course, if A is finite, as we must assume is the case, we should really form the relation $J^{A_1\to A_2}$, where A_2 is another(proper) subset of A. But we need not trouble ourselves with this complication here.)

(10) We may now emphasise the contrast between perceptibility and J-privacy:

(i) From (8) above, we have that V is J-private from J_A in L_A. But V' is directly perceptible to J_A.

(ii) Similarly, in the case of J_B, which we may take to be in all relevant respects like J_A, V is J-private from J_B in L_B. But again, V' is directly perceptible to J_B.

(iii) By Theorem 3, (or by Theorem 2), A is J-private from J_B in L_{B_1}. And A' is indirectly perceptible to J_B; as of course may be other sequences whose members are taken from M but not from V.

(iv) By Theorem 3, A is J-private from J_A in L_{A_1}. And A' is indirectly perceptible to J_A in L_{A_1}.

(v) But A is not J-private from J_A in L_A - by Theorem 1; (this is using the null interpretation for L_A, where the recorded sequence is the recording sequence: A'. Note that in the case of the secondary J-language L_{A_1},A', the recorded sequence, is not the same as A'_1, the recording sequence.).

(vi) We may also say, from (8) above and Theorem 2, (4.32,(11a)), that A is J-private <u>to</u> J_A.

4.42 I will now make some more general remarks about the J-linguistic treatment of the monistic hypothesis that visual imagings (and mental events generally) are neural events.

(1) We make the plausible assumption that we could describe a certain neural structure as a J-structure J_A (the way I have phrased it in 4.41, J_A is a neural structure in a particular subject) for which A' is a sequence of neural events occurring in J_A, and these neural events can be construed as wffs of the J-language L_A interpreted so that they J-designate visible events. It seems to me that in principle this J-linguistic treatment of neural events is unobjectionable. And for the neural structure J_A, the monistic hypothesis that A is a subset of M is quite straightforward and unobjectionable. But we hypothesise that A' is a sequence of visual imagings : - essentially that those events which through introspection we may judge to occur in us when we see things, are neural events.

The notion of J-designation is at least *prima facie* applicable to visual imagings whether or not these are neural events. The real difficulty is of course overcoming the intuitive resistance to the view that they are neural events. But, as anticipated in 4.2, we shall see that the supposition that if A' were a sequence of visual imagings A could not be a subset of M, is unjustified, and tends to be based on a simple confusion. Moreover, if we avoid this confusion things are more or less as we should expect them to be if visual imagings were neural events.

(2) Through the notion of J-privacy we can maintain a sense in which an observing event - a J-designating event - may only partially capture features of the observed event - the J-designated event. But we can do this without invoking such notions as that of the "thing-in-itself". There is no room for something which is additional to the J-designated event but which is in principle imperceptible. An event J-designating a cat jumping off a chair is distinct from the latter event. But there is not also a "cat-in-itself". What may happen to promote the "thing-in-itself" confusion is that ironically we may let the realist distinction lapse: in thinking that there is a "cat-event" we

don't perceive as well as a "cat-event" we do perceive, what may happen is that what we take as the cat-event is really the J-desiignating event - we may look at the cat and think 'there must be something out there which we are not perceiving,as well as the cat we perceive' as a confused symptom of the fact that there is something out there which is additional to the imaging occurring in us. But it is the something out there which we perceive (though we can be introspectively aware of the imaging). A more deliberated confusion having a similar effect may involve the view that during perception a subject is directly aware of or indeed perceives sense-data - leaving an "external reality" unperceived.

But we can and must make the realist distinction in such a way that the view that there are imperceptible "things-in-themselves" involves an incoherent misuse of the concept of perception. I emphasise this point for this reason: although the confusion concerning "things-in-themselves" may be thought a rather primitive one, the sort of thinking involved tends to survive persistently in various guises - in particular, there is the persistant twin confusion which sustains the intuitive resistance to the view that we could perceive mental events. These twin confusions are aided and abetted by the mistaken division between a subject and the mental events he is aware of.[7]

For the moment we must note then that through the J-linguistic concepts of J-privacy and accessibility we can express the possibility that a J-desiignating event may have quite different properties from those of the J-desiignated event which it maps; but without being lured into the infelicitous expression of this possibility through the concept of perception. So we can easily accommodate the fact that our imagings/neural perceptual events may be quite different from and only partially capture their J-desiignata - without supposing that we don't perceive the J-desiignata.

(3) There is an asymmetry of access involved in the dualist's idea that a subject somehow has privileged access to his imagings (and his mental events generally): the symmetry with which subjects C and D may have access to a physical event is broken, according to the dualist, when we consider the access C and D may have to C's mental events.

Now, there is an asymmetry which is present in the J-linguistic treatment, but which is compatible with and may be expected to arise within a monistic domain. We see from 4.41,(10) that we have the asymmetry:

A is not J-private from J_A in L_A (i.e., A is univocally totally accessible to J_A in L_A); but A is J-private from J_B in L_B.

Clearly <u>this</u> asymmetry may arise within a monistic domain M containing A and B. We can illuminate this asymmetry by noting the connection between J-privacy and identity: where L_K is a J-language for M, as a result of Theorem 1,

$(M' = K') \supset M$ is univocally totally accessible to J_K in L_K.

Suppose we hypothesise the bi-implication:

$(M' = K')$ iff M is univocally totally accessible to J_K in L_K.

On the hypothesis that the bi-implication obtains, J-privacy would arise wherever the recording sequence K' is distinct from the recorded sequence M': essentially, where the "observer" and the "observed" are <u>two</u>. I can give no proof for this hypothesis that the bi-implication obtains, but it seems

plausible. And if the bi-implication does obtain, then the asymmetry regarding J-linguistic access boils down to the difference between the relation between something and itself, and the relation between something and something else.

(4) Suppose J_A is scanning A' and a_n occurs.[9] Then not only will a_n occur, it will also cause $a(a_n)$ to occur. And then we can say that J_A (indirectly) perceives a_n. But we cannot say that J_A perceives a_n just in virtue of the occurrence of a_n - nor can we give any sense to J_A's being <u>aware</u> of a_n as it occurs.

But suppose J_A is part of the richer structure J_A^+, rich enough to sustain events which may be described as introspective: suppose J_A^+ is able to introspectively compare a_n and $a(a_n)$. (Though, as emphasised earlier, we do not require J_A^+ to be introspectively aware of the J-designating events in order to perceive the J-designated events.) We know that in this case J_A^+ will find that a_n and $a(a_n)$ are different, though isomorphic.[10] Indeed, the J-relation will obtain between a_n and $a(a_n)$, not identity (cf. 4.32,(16)).

But J_A^+ in confusion may suppose that because $a(a_n)$ is different from a_n he cannot be perceiving a_n. J_A^+ may in confusion regard $a(a_n)$ as the J-designated event; in which case, since $a(a_n) \neq a_n$, he may wrongly deduce that a_n cannot be the J-designated event. But this confusion is not unnatural, since (crudely) normally if J_A^+ introspectively compares the events occuring in him when he judges perceived events to be similar, he finds the former events similar.

And suppose J_B^+ perceives a_n, which J_B^+ "expects" to be like b_n; J_B^+ might confusedly compare (introspectively) $b(a_n)$ with b_n, and finding these to be different, wrongly deduce that he could not be perceiving something like b_n.

Events in J_A^+ which may be describable as thinking, feeling pain, and so on - which are not in an obvious way interpretable as J-designatory" - would again be quite different from events in J_A^+ which J-designated them. And again the confusion to be avoided is for J_A^+ to suppose that since the event J-designating a pain-event is different from the pain event, he could not be perceiving a pain event.

It is this sort of confusion which strikes me as underlying the intuitive resistance to the view that if we could perceive certain neural events, we would then be perceiving events such as visual imagings, trains of thought, and so on.

(5) Clearly much more needs to be said about the notion of introspection involved above. But I will confine myself here to the following brief comments.

(a) I must stress again that we can and must accept that events such as imagings, experiences of pain, trains of thought, etc., all occur without the need to posit a subject which "apprehends" them. There may be some resistance to the view that we could feel pain as we do without there being a subject to be "aware" of the pain. It might be thought that if we are not "aware" of pain-events then we must be led to some sort of automatism in which mental experience becomes a "blank", and in which we could not be said to <u>feel</u> pain. But the sort of events which we insist occur without a subject being aware of them, are what we may describe as <u>feelings</u> of pain, trains of thought, imagings and so on. A segment of our conscious experience does not have to

include these events plus the awareness of them by a subject. Suppose someone is reacting to a complex train of external events, during which there occur in him imagings, pain-events, a train of thought - but he is concentrating so hard he does not at the time think about these mental events : he does not think 'Ah! I'm in pain', 'A visual imaging is occurring in me'. The imagings, and so on, occurred nevertheless - his mental experience was not "blank". What may give rise to the illusion that in addition to the "naked" occurrence of these events, they must be "apprehended" by a subject, is that these events may be accompanied by events which we would describe as thoughts about them : a pain-event may be accompanied by the thought 'Ah! I'm feeling pain' - and the latter may be described as introspective. The occurrence of such accompanying thoughts is all we need in order to construe introspective awareness of imagings and so on.[12]

(b) Suppose a train of thought t occurs in someone. We should expect that, like any physical event, the occurrence of t is no more and no less than its occurrence. But what may contribute to the dualist's confusion is that when he perceives physical events he can tell "what they're physically like" - but the occurrence just of t must not be expected to be accompanied by an "awareness of what it's physically like"; we can only be expected to describe an event in physical terms if we can perceive it (directly or indirectly); but in order for us to be able to perceive a train of thought t, there have to occur in us J-designations of t. So, as with any physical event, just the occurrence of t must not be expected to "reveal to a subject" where t is occurring, for example, or what colours we would see if we saw t, and so on. Thus the events in a sufficiently rich J_A^+ are peculiar in this respect: loosely speaking, for J_A^+ to have evidence of the occurrence of most (physical) events, J_A^+ must perceive them directly or indirectly. (To simplify the point I'm trying to make I'm not counting evidence from other observers - it might simplify matters if we considered J_A^+ as a solitary observer.) But we can see that there is a certain sort of physical event for whose occurrence a sufficiently rich J_A^+ may have evidence of a sort even though contingently J_A^+ cannot perceive them: events in J_A^+.

(c) But, what is perhaps more important for my present purpose, if we cannot allow a sense in which a person can be introspectively aware of his mental events, the dualist cannot even begin to justify ontological dualism. While if we can allow introspective awareness of mental events, we can undermine the justification for ontological dualism: the dualist cannot easily justify the view that the events he is introspectively aware of are radically different from the events he perceives. Certainly, if the dualist is looking at his own or someone else's brain, and has a visual imaging as a result of which he may say 'I see something grey', and if he introspectively compares this imaging with, for example, a pain-event (which the neural event is supposed to be), then it is not at all unexpected that the two events should be found introspectively to be quite different. But what grounds has he for supposing that he is not perceiving a neural event/pain-event? What grounds has he for supposing that the J-relation, which obtains between his imagings and events perceived, does not obtain between his imaging when looking at the neural event, and the pain-event? We must remember that it is to be expected in this case that he should be able to be introspectively aware of the occurrence of events which he may not be perceiving, and which he cannot describe intrinsically in physical terms until he perceives them.[13]

4.43 We have already in the case of L_{A_1} considered the possibility of syntactical reference: wffs of L_{A_1} may be designated by wffs of L_{A_1}; and there is no logical difficulty in envisaging this possibility. I will now consider the possibility of self-reference in J-languages.

Clearly self-reference is possible in any J-language on the null interpretation, where every wff is interpreted as designating itself. But it is inconceivable that the null function is the only j-function which a J-structure can be taken as operating. In this sense we can regard the null-function as inoperable: a J-structure J_K could not be said to generate K' according to and only according to the rule '$k_n = k_n$', even though it would always be the case that $k_n = k_n$.

What we will do here is find a way in which a wff can be interpreted as

designating itself on a non-trivial interpretation - not the null interpretation, but an interpretation given by an operable j-function. But we will find, as we might expect, that such a self-designating wff could not be generated by the J-structure.

We will consider the possibility of non-trivial self-reference in the J-language L_{A_1}, so that the issue may immediately be put in terms of autocerebroscopy.[14]

(1) More or less as before (in 4.41):

M is a set of physical events.

V is a (proper) subset of M comprising just visible events.

A' is generated by J_A, which is "normally" veridical in L_A for V.

A is a (proper) subset of M; but A and V are disjoint (i.e. they have no common elements).

(2) $\overset{A\to V}{J}(a_n, v_n)$; $a(v_n) = a_n$.

(3) J_X is an amplifying J-structure and X is a (proper) subset of V such that $\overset{X\to A}{J}(x_n, a_n)$: L_X is a J-language for A. So since $\overset{A_1\to X}{J}(a(x_n), x_n)$ where $\overset{A_1\to X}{J}$ is a restriction of $\overset{A\to V}{J}$, A_1 being a proper subset of A,

$\overset{A_1\to A_2}{J}(a(x_n), a_n)$; where A_2 is another proper subset of A. (We are assuming A is finite.)

Thus $\overset{A_1\to A_2}{J}$ gives a secondary domain of interpretation for wffs of L_{A_1}, which is a secondary J-language for A_2.

(4) I will now describe J_X and the sequence X'.

The J-language L_X is unusual in comparison with the J-languages we have been considering, in this respect: the wff-events of L_X will of course J-designate in L_X ; but in addition, the wff-events of L_X will be interpretable as designating in a formal language which is a specialised form of English, S-English. The reason this device is used is so that we may have a more familiar index of J-designation in L_X :- it may be compared to the device of translating photo-electronic events in a television camera onto a screen so that the J-designatum of a photo-electronic event is depicted on the screen.

Part of J_X is a screen; and on the screen may appear a (finite) designation in S-English of an event from A. Designation of members of A in S-English is achieved as follows (- I will use double quotation marks for meta-S-English use, since I will use single quotation marks in S-English):

(i) We can construct in S-English expressions which designate members of V. But except for one case which we deal with later, we need not specify how these designators are constructed - though we must assume there is a formal specification for the construction of these designators. I will use the notation:

$\underline{v}^i$ is an <u>expression</u> in S-English which designates (in S-English) v^i (a member of V).

(ii) By using the signs "a","(",")", we can use the facility described as available in (i) in order to designate members of A in S-English, as follows:

If $\underline{v}^i$ is an expression in S-English designating v^i (a member of V), then $\ulcorner a(\underline{v}^i)\urcorner$ designates $a(v^i)$ in S-English.[15]

$X \rightarrow A$

Then X' is the sequence of visible events on the screen of J_X. And J, operated by J_X (which is assumed to be veridical in L_X), is a bijection - this is guaranteed by the fullness of the descriptions of members of V in S-English.

So, for each member of A there is a corresponding member of X such that the latter, when interpreted as a series of signs in S-English designates the member of A in S-English, and when interpreted as a wff of L_X J-designates that member of A in L_X:

x^i J-designates a in L_X iff x^i designates a^j in S-English; i.e.,

$J^{X \rightarrow A}(x^i, a^j)$ iff x^i designates a^j in S-English.

(5) Now, the sequences V', A' and X' are to be generated in virtue of the causal relationships between the generating structures, such that v_n causes a_n which causes x_n. What causes v_n need not concern us here: it is enough to assume that V' is generated. So when J_A is scanning V' we have:

V'	A'	X'	(I'm using the obvious device of re-
•	•	•	presenting the temporal order of the
•	•	•	wff-events by means of the vertical
v^i →	$a(v^i)$ →	$x(a(v^i))$	lists. The arrows represent the
•	•	•	causal relationships.)
•	•	•	

(6) The event $x(a(v^i))$ is the <u>appearance on the screen</u> of $\ulcorner a(\underline{v}^i)\urcorner$ (- the latter expression designates in S-English the J-designatum of $x(a(v^i))$). To emphasise this I will use the following notation:

$x(a(v^i)) = [a(\underline{v}^i)]$, where "$[a(\underline{v}^i)]$" is read as:

"The event occurring on the screen when $\ulcorner a(\underline{v}^i)\urcorner$ appears on it". That's to say, when $[a(\underline{v}^i)]$ occurs, "a( )" appears on the screen with the expression $\underline{v}^i$ inside the brackets. And in general, if E is an expression of S-English, then [E] is the appearance of E on the screen. When the screen is blank we will sat that[ ]is occurring.

(7) We now introduce the term "Screen" of S-English. This term is used in designations of visible events occurring on the screen, by means of the following construction:

Let F be an expression of S-English which designates an expression of S-English. To enable the construction of such an expression F we allow the use of single quotation marks in S-English. Then $\ulcorner$ScreenF$\urcorner$ (i.e. "Screen" followed by F) designates in S-English the visible event occurring on the screen when F appears on it. And the screen when blank is designated in S-English by "Screen". Thus:

"Screen" designates in S-English the event: [ ]

$\ulcorner$ScreenF$\urcorner$ " [F]

$\ulcorner$Screen'a($\underline{v}^i$)'$\urcorner$ " $[a(\underline{v}^i)]$

So, suppose J_A had been scanning <u>the screen</u> (i.e. X') when $[a(\underline{v}^i)]$ occurred; we would then have:

V'	A'	X'
•	•	•
•	•	•
$[a(\underline{v}^i)]$ →	$a([a(\underline{v}^i)])$ →	$[a(\text{Screen}'a(\underline{v}^i)'))]$
•	•	•

<u>Explanation</u>: Let $[a(\underline{v}^i)] = v^j$; (although $[a(\underline{v}^i)]$ is a member of X (= $x(a(v^i))$), we must remember that X is a subset of V: and here X' is

part of V').

Then we would have:

$$\begin{array}{ccccc} V' & & A' & & X' \\ \cdot & & \cdot & & \cdot \\ v^j & \rightarrow & a(v^j) & \rightarrow & [a(\underline{v}^j)] \\ \cdot & & \cdot & & \cdot \end{array}$$

Now $a(v^j) = a([a(\underline{v}^i)])$, and since $\underline{v}^j = \ulcorner \text{Screen'}a(\underline{v}^i)\text{'} \urcorner$ ($\underline{v}^j$ being an expression of S-English which designates v^j), we have:

$$[a(\underline{v}^j)] = [a(\text{Screen'}a(\underline{v}^i)\text{'})] .$$

(8) In order to simplify slightly the expressions involved, I will modify S-English to S'-English. S'-English is exactly like S-English, except that:

$[a(\underline{v}^i)]$ designates $a(v^i)$ in S-English iff

$[\underline{v}^i]$ designates $a(v^i)$ in S'-English. But $\underline{v}^i$ still designates v^i in S'-English.

In other words, if we discard the signs "a","(",")" from expressions in S-English, we obtain their equivalents in S'-English. And if expressions on the screen of J_x are now in S'-English instead of S-English, we get the simpler:

$$\begin{array}{ccccc} V' & & A' & & X' \\ \cdot & & \cdot & & \cdot \\ v^i & \rightarrow & a(v^i) & \rightarrow & [\underline{v}^i] \\ \cdot & & \cdot & & \cdot \end{array}$$

And:

$$\begin{array}{ccccc} \cdot & & \cdot & & \cdot \\ [\underline{v}^i] & \rightarrow & a([\underline{v}^i]) & \rightarrow & [\text{Screen'}\underline{v}^i\text{'}] \\ \cdot & & \cdot & & \cdot \end{array}$$

(9) We now introduce a further sign for S'-English: "N".

"N" is used in S'-English as the sign for the norm function as described by Smullyan (cf. Smullyan, pp. 64-6). By the norm of an expression of S'-English we mean the expression followed by its own quotation.

We already have in S'-English the device handed down from S-English whereby we may designate expressions of S'-English by expressions of S'-English through the use of single quotation marks. We now have a second device for the designation in S'-English of expressions of S'-English:

If F is an expression of S'-English which designates the expression E of S'-English, then $\ulcorner NF \urcorner$ designates in S'-English the norm of E: E followed by its quotation. Thus $\ulcorner N'E' \urcorner$ designates $\ulcorner E'E' \urcorner$ in S'-English. So $\ulcorner N'E' \urcorner$ and

$\ulcorner$'E'E''$\urcorner$ co-designate in S'-English. And so, for example, $\ulcorner$ScreenN'E'$\urcorner$ and $\ulcorner$Screen'E'E''$\urcorner$ co-designate in S'-English. We should note also that the expression "N'N'" designates itself.

Now, J_χ is such that, where an expression of the form $\ulcorner$'E'E''$\urcorner$ would appear on the screen, $\ulcorner$N'E'$\urcorner$ appears instead.

(10) We now consider the event $x^{\sigma'}$ = [ScreenN'ScreenN']

And we find that $x^{\sigma'}$ designates in L_χ $a(x^{\sigma'})$.

Explanation: In (4) above we stipulated that

x^i J-designates a^j in L_χ iff x^i designates a^j in S-English.

Adapted for S'-English this becomes:

x^i J-designates a^j in L_χ iff x^i designates a^j in S'-English.

So, to find which member of A is J-designated in L_χ by $x^{\sigma'}$ we simply find which member of A is designated by $x^{\sigma'}$ in S'-English.

Let $x^{\sigma'}$ = $[\underline{v}^h]$. We have stipulated in (8) above that

$[\underline{v}^h]$ designates $a(v^h)$ in S'-English ...(i)

where

$\underline{v}^h$ designates v^h in S'-English ...(ii)

Suppose $\underline{v}^h$ = $\ulcorner$ScreenF$\urcorner$; then $\underline{v}^h$ designates in S'-English the appearance of F on the screen - i.e. $\underline{v}^h$ designates [F] in S'-English. But in this case $\underline{v}^h$ = "ScreenN'ScreenN'", which designates in S'-English the appearance of $\underline{v}^h$ on the screen. So $\underline{v}^h$ designates $[\underline{v}^h]$ in S'-English. So, by (ii), $v^h = [\underline{v}^h] = x^{\sigma'}$. So, by (i), $x^{\sigma'}$ designates $a(x^{\sigma'})$ in S'-English. So $x^{\sigma'}$ J-designates $a(x^{\sigma'})$ in L_χ.

(11) If we had stuck to S-English throughout without adopting S'-English, we would instead introduce the terms "D" and "y" into S-English, "D" being used as the sign for the diagonal function as described by Smullyan (cf, Smullyan, p.67n). The diagonalisation of an expression E of S-English would be the result of substituting the quotation of E for all occurrences of "y" in E. So $\ulcorner$D'Fy'$\urcorner$ and $\ulcorner$'F'Fy''$\urcorner$ co-designate in S-English - likewise $\ulcorner$D'(Fy)'$\urcorner$ and $\ulcorner$'(F'(Fy)')'$\urcorner$.

So, for example, ⌜a(ScreenD'Fy')⌝ and ⌜a(Screen'F'Fy'')⌝ co-designate in S-English. And J_x would then be such that where an expression of the form ⌜a(Screen'F'Fy'')⌝ would appear on the screen, ⌜a(ScreenD'Fy')⌝ appears instead. 16b

(The reason why we have to use the diagonal function in S-English is that here the quotation is embedded in an expression rather than following on as in S'-English.)

We would then consider the event x^G = [a(ScreenD'a(ScreenDy)')], and again we find that x^G would J-designate in L_x the event $a(x^G)$, since [$a(\underline{v}^i)$] designates $a(v^i)$ in S-English, and here "ScreenD'a(ScreenDy)'" designates x^G in S-English - since it designates the appearance on the screen of 'a(Screen D'a(ScreenDy)')'.

Although x^G is more cumbersome to work up to than $x^{G'}$, we would avoid the rather over-economical feature of J_X operating in S'-English, where [$\underline{v}^i$] designates $a(v^i)$ while $\underline{v}^i$ desigantes v^i.

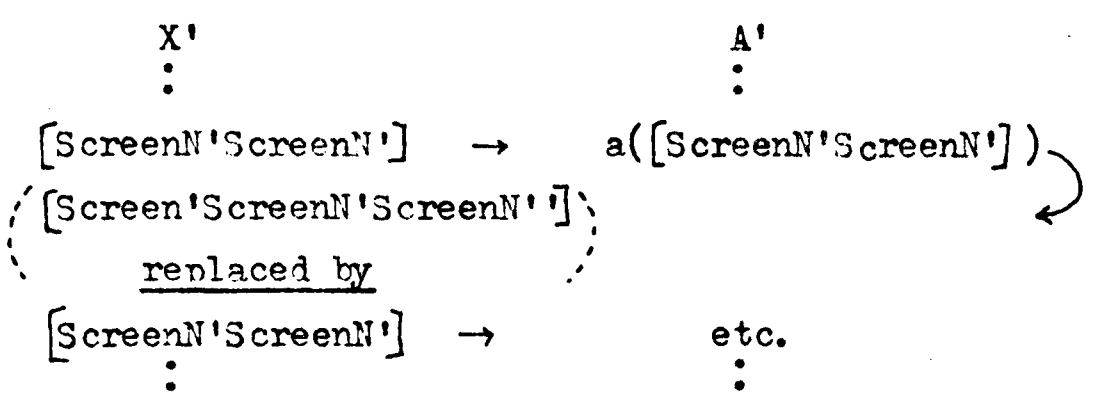
(12) From (11) we have, in the case that J_X is operating with S-English, ${}^{X\to A}J(x^G, a(x^G))$; but ${}^{A_1\to X}J(a(x^G), x^G)$, so ${}^{A_1\to A_2}J(a(x^G), a(x^G))$. And similarly, if J_X is operating with S'-English, then ${}^{A_1\to A_2}J(a(x^{G'}), a(x^{G'}))$.

So, if we let $a(x^G) = a^G$, a^G <u>J-designates itself in L_{A_1}</u> in the S-English case; and $a^{G'}$ ($= a(x^{G'})$) <u>J-designates itself in L_{A_1}</u> for the S'-English case.

(13) As may be expected, on the assumption that J_A and J_X are veridical, neither a_n^G nor $a_n^{G'}$ are derivable. To show this I will concentrate on the S'-English case and $a^{G'}$ and $x^{G'}$. But similar considerations apply to a^G and x^G.

We can see immediately that neither $a_m^{G'}$ nor $x_n^{G'}$ would be generated if J_A is scanning the screen and the screen is blank when the scanning starts: in this case successive members of X' would be [ ], [Screen], [N'Screen'] (replacing [Screen'Screen']), [Screen'N'Screen''], and so on. (It may be thought that we could simply arrange J_x so that if "Screen'Screen'" would appear on the screen, this is replaced not by "N'Screen'" but by "ScreenN'ScreenN'". But in this case [ScreenN'ScreenN'] would be false in L_x if generated in this manner.)

But we can show that $a_m^{G'}$ and $x_n^{G'}$ would not be derivable for any sequences A' and X', providing J_A and J_X are veridical in L_A and L_X respectively. Suppose X' contains $x_n^{G'}$; then we would have:



And suppose X' and A' are finite, which they are ex hypothesi if $a_m^{G'}$ and $x_n^{G'}$ are derivable. We also have to assume that J_A and J_X are operating recursively, in roughly this sense: if J_K is a physically realised J-structure operating its j-function recursively, then w_n , the "input" for the j-function, causally precedes the "output" k_n .

Now, if A' and X' are finite, there must be first elements of both. So:

(a) Suppose $a_m^{G'}$ is the first element of A', i.e., $a_m^{G'} = a_1^{G'}$. In this case, either $a_1^{G'}$ is false' in L_A , or $x_1^{G'}$ is the first element of X'. But then either $x_1^{G'}$ is false' in L_X , or $a_1^{G'}$ is not the first element of A'. Unless, that is, $a_1^{G'}$ does not causally precede $x_1^{G'}$ and vice versa, in which case J_A and J_X are not operating recursively.

(b) Suppose then that $a_m^{G'}$ is not the first element of A'. But if A' is finite, $a^{G'}$ must be closed in A' at some earliest position, say e. But then either $a_e^{G'}$ is false' in L_A , or its J-designatum $x_n^{G'}$ is false' in L_X , or $a_e^{G'}$ is not the earliest closure of $a^{G'}$.[17]

So either $a_m^{G'}$ and $x_n^{G'}$ are not derivable, or either J_A or J_X or both are inconsistent (and inveridical).

But suppose that $a_1^{G'}$ were generated as the first element of A'. In this hypothetical case $a_1^{G'}$ would be true' in L_{A_1}, since $J^{A_1 \to A_2}(a_1^{G'}, a_1^{G'})$. So $a_1^{G'}$ is true^2 in L_{A_1}, though it is not derivable (cf. 4.32 (15a) for the definition of truth^2).

The same argument applies to the case of a^G and S-English.

(14) So we have

Theorem 4: If J_X is veridical in L_X, then J_A is either incomplete[2] in L_{A_1} for A_2, or inconsistent in L_A for V.

It may be thought that this result applies only to the particular amplifying structure described. But it applies also to any other amplifying structure J_Y for A, in which the norm/diagonal function facility is available, since if J_Y operated $\overset{Y\to A}{J}$, then $\overset{Y\to A}{J}(y(a_n),a_n)$ iff $\overset{X\to A}{J}(x(a_n),a_n)$ - if J_Y and J_X are both veridical in their respective L_Y and L_X.

The argument of (13) above applies to any wff of L_{A_1} which J-designated itself in L_{A_1} for A_2, not just to a^G and $a^{G'}$: in particular, the arguments at (a) and (b) in (13) would apply to any such wff and its J-designatum in L_A for V, in place of $a^{G'}$ and $x^{G'}$. So we have:

Theorem 5: If a wff J-designates itself in L_{A_1} for A_2, it is not derivable by J_A, provided J_A is veridical.

(15) The following informal comments may help to bring out some salient features of the results obtained.

(a) Suppose we represent the j-function for a J-language L_K as follows: if $j(k^i) = w^h$, we represent this as: '$k^i \overset{j}{\leftarrow} w^h$'. This of course represents that k^i J-designates w^h in L_K. Now, if J_K is a physically realised J-structure operating j, there is a causal relation which we may represent by '$\overset{c}{\leftarrow}$', which in a sense physically realises the j-function:

$k^i \overset{j}{\leftarrow} w^h$ iff $k^i \overset{c}{\leftarrow} w^h$.

Now, it is intuitively clear that in as much as we can conceive that k^i_n is generated we must suppose it stands in the J-relation to something else, and stands in the causal relation to something else:

$k^i \overset{j}{\leftarrow}$; $k^i_n \overset{c}{\leftarrow}$.

If we consider the null function, we find that $k^i \supset n$; and so the null function cannot be physically realised in this sense: it does not determine a recursive relation between k^i and something else, which may be physically realised by a causal relation. In other words, the null function cannot be physically realised by a causal relationship whereby k^i_n can be generated as part of a "causal chain".

Now, for J_A and J_X we suppose there are operable j-functions which may be physically realised by causal relations:

$$x^h \xleftarrow{j} a^i \ , \ x^h_n \xleftarrow{c} a^i_n \ ; \quad a^i \xleftarrow{j} v^h \ , \ a^i_n \xleftarrow{c} v^h_n \ .$$

Here x^h J-designates a^i in L_X. Of course, a^i may designate a member of X, but, barring the exceptional case of the pair a^G and x^G, if x^h J-designates a^i in L_X, a^i will not J-designate x^h in L_A - the arrow of designation normally only goes one way, if at all, between a pair, and likewise the arrow of causation.

But, having established operable j-functions, we form the case of a^G and x^G, where we have:

$$a^G \xleftrightarrow{j} x^G \quad \text{and} \quad a^G_m \xleftrightarrow{c} x^G_n \ .$$

(This is simplifying, because the j-function for L_X is different from the j-function for L_A : $a^G \xrightarrow{j} x^G$ is different from $a^G \xleftarrow{j} x^G$. But we picture the main point that in their respective J-languages they J-designate each other.)

So in effect we have $a^G \supset j$ and $a^G_m \supset c$ (and $x^G \supset j$ and $x^G_n \supset c$). So we can see intuitively that a^G and x^G must lie outside the chain of generation, and are in a sense "causally independent" of any causal chain. This somewhat resembles the way the Gödel sentence for a suitable system lies outside a "chain of derivability".[18]

(b) A little more may be usefully said about the norm/diagonal function facility in the amplifying structure J_X.

First we must note that if this facility is not available, then what I have called syntactical reference is still possible by means of a weakened J_X But it is only if J_X is strengthened by the availability of the norm/diagonal function that a self-designating wff can be constructed. I will not support this rigorously, but offer the following brief comments.

We must note the contrast between the quotation device for designating, and the norm/diagonal function device.

Suppose that d_1 is a function from expressions in a system S to expressions in S, such that the members of the range (the values) of the function are designations in S for members of the domain of the function. And let d_1 be such that, if E is an expression of S, $d_1(E) = \ulcorner 'E' \urcorner$; $d_1(E)$ is the value

of the function d_1 for the argument (domain member) E. So $d_1(E)$ designates E. We can see intuitively that given this rule for the construction of designators, a designator $d_1(E)$ must always be different from the expression E which it designates : for every E, $d_1(E) \neq E$.

We may usefully call designation as prescribed by d_1 , "first order" designation. Note that of course the pair of quotation marks does not work in S as a sign for the function d_1 - the quotation marks simply alter strings of S as prescribed by the function d_1. But there could be a sign in S which works as the sign for the function d_1 , e.g., the sign "Q" such that $\ulcorner QE = \text{'}E\text{'}\urcorner$ is true in S for every E (i.e. $\ulcorner QE \urcorner$ does the work of "$d_1(E)$"). Then $\ulcorner QE \urcorner$ designates in the system the expression E. But again, $\ulcorner QE \urcorner$ cannot designate itself.

Now consider the "second order" function d_2 from expressions to expressions: $d_2(E) = \ulcorner Ed_1(E) \urcorner$. So d_2 is the norm function, and if there is a sign "N" in S for the norm function, we can have "second order" designation: $\ulcorner NE \urcorner$ designates $\ulcorner Ed_1(E) \urcorner$ in S; and where E = 'N' , NE = $\ulcorner NE \urcorner$.

Now, if only "first order" designation is available for S, prescribed through a function d, then at least in the following case we can show that non-trivial self-designation is impossible. (By trivial self-designation I mean the case where every expression of S designates itself). Suppose that d is such that for each expression E of S, d prescribes one and only one expression of S as the designator of E. Clearly in this case for every E d(E) must be different from E - barring the trivial case where d(E) = E for every E. d_1 is such a function, since it prescribes one and only one designator - namely $d_1(E)$ - for each expression E of S.[19]

Now, the availability of suitable "second order" designation in S'/S-English through the norm/diagonal function allows the construction of a wff x^G which J-designates $a(x^G)$. With only "first order" designation available in S'/S-English, it becomes impossible to construct such an x^G , since to permit its construction we have to rely on the self-referential capacity of S'/S-English.

We can easily produce an example which resembles the case of J_X without the norm/diagonal facility. Suppose J_C is a television camera and J_S a television screen; J_S can be taken as the amplifying structure for J_C. L_C is a J-language for V (visible events) and L_S is a J-language for C, such that:

s^i (wff of L_S) "depicts" v^j iff s^i J-designates $c(v^j)$ in L_S.

In an ordinary J_S we only have "first order" designation involved in the "depicting", and we could not envisage the possibility of a self-J-designating wff in this case. (Imagine the camera scanning the screen: a picture S^i on the screen could not depict the screen showing s^i - if s^i depicted n screens it would not depict the screen with n screens depicted on it[20].)

(16) The case of J_A scanning the screen of J_X is clearly a formalisation of certain features of an autocerebroscopic situation. As a result we can from Theorem 5 confirm the expectation that in an autocerebroscopic situation there could not occur an imaging/neural event which J-designated itself.

Now, in view of the symptoms of the pathology of self-reference which the autocerebroscopic situation shows, it may be wondered whether the results expressed in Theorems 4 and 5 are at all relevant to arguments such as that of Lucas, which is directed towards the view that Gödel's Theorem rules out a "mechanistic" conception of the operation of the human mind. Since the autocerebrosconic case centres around the issue of perception rather than the sort of "mental" operation discussed in Lucas's argument, it is clear that there can only be an indirect connection between the two issues. But in order to put what connection there is in some perspective, I will comment briefly and superficially on Lucas's argument.

Lucas says:

> Now any mechanical model of the mind must include a mechanism which can enunciate truths of arithmetic, because this is something which minds can do: in fact, it is easy to produce mechanical models which will in many respects produce truths of arithmetic far better than human beings can. But in this one respect they cannot do so well: in that for every machine there is a truth which it cannot produce as being true, but which a mind can. This shows that a machine cannot be a complete and adequate model of the mind. It cannot do <u>everything</u> that a mind can do, since however much it can do, there is always something which it cannot do, and a mind can.
>
> ((Lucas, p.47.))

This sort of argument supposes that while there must be an arithmetical blindspot for a machine which is supposed to enunciate truths of arithmetic, there is no corresponding blindspot for a human mind. Clearly this sort of argument is very vulnerable to the objection that it begs the question.[21] Presumably proponents of such arguments find it difficult to conceive what arithmetical statement would be such that in principle they could not verify it unless it was false.

If Lucas's argument is to be taken seriously, the sense in which there is supposedly no blindspot for the human mind must be the sense in which there is a blindspot for a Turing machine. But the issue tends to be clouded by this factor: there is a tendency in Lucas's arguments to imbue the sense with which 'proof', 'prove', 'derivable', etc. are applied to formal systems and Turing machines, with the colloquial sense with which we can be said to verify, prove, derive and so on. But this is a dangerous tendency, since derivability in a formal system is essentially a syntactical property. Thus a Turing machine <u>may</u> in its production of outputs be interpretable as "proving", "deriving" in the colloquial sense, but by no means necessarily so.

Now, although the blindspot for a formal system or a Turing machine can always be interpreted as an <u>arithmetical</u> limitation - whether or not the system or machine is of primary interest with regard to the arithmetical interpretation of its formulae or outputs - it may not be immediately obvious as an arithmetical limitation.

The limitation or blindspot is more generally construable as this: if the derived formulae of a system are true on a certain interpretation, then if the system is sufficiently rich there is a formula of the system which is true on that interpretation if and only if it isn't derivable in the system. Suppose P(s) is a syntactic relation such that P(s) obtains if and only if the formula s is derivable in S. If S is rich enough to contain formulae interpretable as saying that P(s) obtains (for suitable s's), then a formula t, interpretable as saying that $\sim P(t)$, is true if and only if it is not derivable in S. But providing the system S is finitary and formulae are derived by recursive procedures, we can always "arithmetise" S: there will always be constructable

a Gödel correspondence for the system (a unique assignment of numbers to syntactical entities of S), and an arithmetical relation A such that if $\bar{s}$ is the Gödel number assigned to s, $A(\bar{s})$ iff P(s). Then if the formulae of the system are interpretable as arithmetical statements, a formula g interpretable as the arithmetical statement '$\sim A(\bar{g})$' would be true in S if and only if it wasn't derivable in S. But notice that whether or not the intended interpretation of S interprets formulae as arithmetical statements, provided a formula is intperpretable as saying P(s), ipso facto it is interpretable as saying $A(\bar{s})$, since the formula is true in S if and only if $A(\bar{s})$. In other words, the intended interpretation of S does not have to be arithmetical in order for us to conceive the limitation as an arithmetical limitation.

Now, if we transpose this sort of consideration to the case of physically realised Turing machines, then formal derivability becomes physical "outputtability". And if we know that a physically realised Turing machine M_P is such that it "models" - is a physical realisation of - the abstract machine M_A (or if we know "outputtability" in M is designed to mirror derivability in a certain formal system), then the question of whether a certain output s can occur in M is not one which requires an investigation of the causal mechanisms in M_P. For ex hypothesi the causal mechanisms - whatever they're like - are harnessed in M_P so as to mirror the abstract specifications in M_A. As a result M_P is no more difficult to arithmetise than is a formal system, and we can construct a relation A such that $A(\bar{s})$ obtains if and only if $\bar{s}$ is the Gödel number of an "outputtable" formula. This of course enables the prediction of what will occur as outputs of M_P. And the limitation of M_P, its blindspot, amounts to this: M_P will be unable to determine correctly for every s whether or not it will produce s as an output. (Or, in other words, it cannot compute correctly for every s whether $A(\bar{s})$ obtains.) So we can regard the limitation of M_P as an inability to predict correctly everything that will occur in it: since there is an output which can occur if and only if it is predicted by M_P that it won't occur. (This is particularly salient in the presentation of this limitation in terms of the so-called "halting problem" for Turing machines. This is a result of Turing's very

similar to Gödel's but couched in a Turing machine setting; putting it crudely, it involves a formalisation of the question 'Will you output a negative answer to this question?'[22].)

But if we do not know the specifications for a Turing machine M_P (physically realised), then predicting what will occur in M_P, and the parallel problem of arithmetising M_P, may both be extremely difficult, since they could involve an extremely difficult empirical investigation of M_P. And if a human brain could be regarded as a physically realised Turing machine (in that its operations would be regarded as determined in the same sense as a physically realised Turing machine's would be determined), then the corresponding blindspot would be a limitation in a brain's ability to predict what can occur in it. Put in this way the probability of there being such a limitation is not in the least striking - since the brain is assumed to have only a finite memory and an upper bound to the number of "computational steps" which may occur in it; (the result is more striking in the case of an abstract Turing machine endowed with an infinite memory and allowed any finite number of steps in the production of an output). Then the corresponding arithmetical limitation for a brain would arise in this way: if the brain was arithmetised - which it could be on the assumption that it is in certain respects like a physically realised Turing machine - then there would be an arithmetical relation A such that $A(\bar{s})$ obtained if and only if $\bar{s}$ was the Gödel number of a brain state which could occur. Then our inability to predict correctly for every s whether or not s occurred would manifest itself in an inability to compute correctly whether or not $A(\bar{s})$ obtains for every $\bar{s}$. This would be the corresponding arithmetical limitation, whatever other arithmetical limitations a brain may have.

Naturally Lucas would object to the assumption that a brain could be arithmetised. But he cannot support his objection by the statement that there is no arithmetical blindspot for the human brain - for this really says no more than that the brain could not be arithmetised.

Now, the results of Theorems 4 and 5 may also be transposed into the

possibility of an arithmetical limitation: if we could arithmetise the j-function. (Or more strictly, if we could arithmetise the principle of generation for J_A ; but it is assumed that J_A operates the j-function.) This would involve being able to assign Gödel numbers to members of A and X. Then suppose $\bar{x}^i$ is the Godel number of x^i , and $\bar{a}^j$ is the Gödel number of a^j . Then we would have to suppose there is an arithmetical function g such that $x^i = j(a^j)$ iff $\bar{x}^i = g(\bar{a}^j)$. And then J_A could be interpreted as computing the arithmetical function g: its output a^j_h would be interpretable as a designation of the number $\bar{a}^j$; so it would be interpretable as computing the arithmetical value $\bar{a}^j$ for the input $\bar{x}^i$.[23] But J_A could not compute in every case that $\bar{x}^i = g(\bar{a}^j)$ - it could not establish that $\bar{x}^G = g(\bar{a}^G)$.[24]

As promised this is all rather brief and cursory; but I hope that some idea of the contextual background to the J-linguistic result is indicated.

4.5 Subjective Dualism, Ontological Monism and the Augmented Identity Theory

The account given so far suggests that the ontological dualist's confusion stems from a contrast which he finds difficult to reconcile with monism, but which is in fact bound to occur if we are purely physical structures. We must expect that (a) if we are purely physical, there nevertheless occur events in us for which we cannot at present give an intrinsic physical description - for the simple reason that contingently we cannot at present perceive them, though we can specify their occurrences in relation to processes which we can describe physically. And moreover we must expect that (b) if we could perceive a mental/neural event ϕ/ϕ, then the perceptual event J-designating ϕ/ϕ will to some degree be isomorphic to ϕ/ϕ, but will have to be different from ϕ/ϕ. We can take (a) and (b) together as an expression of the view I call subjective dualism. And the sort of thought-experiment from which the ontological dualist might derive support for his view cannot coherently be made to justify his view, and cannot disturb ontological monism. For it is suggested that the dualist takes an illegitimate and confused step from subjective dualism to ontological dualism.

To give a heuristic idea of subjective dualism and the dualist's illegitimate step, I again offer an analogy - hoping that through the J-linguistic treatment I have given substance and credibility to the analogy.

Suprose that on a television screen is depicted that television screen - we can imagine a picture on the screen in which the television screen is depicted amongst other objects. The image depicting the television will be different from the total image on the screen - though it will be isomorphic to the total picture. (Clearly in this case the isomorphism is particularly obvious. All we can expect from the J-linguistic application to mental/neural events is an isomorphism.) And unless a picture on the screen contains an image depicting the television screen, the picture will not indicate the screen's physical relation to other objects - in particular, its relation to the objects depicted on the screen. And the dualist's confusion can be likened to the confusion involved in supposing that the television screen could never be any of the objects appearing on the screen because none of the objects which could appear on the screen could ever be the same as the total picture on the screen.

4.51 We can now reconsider the main objections to the Identity Theory, objections through which the dualist(-and the unconvinced materialist-) argues that the mental events occurring in him could not be neural events.

We saw in Part 3 that the II-Identity Theory didn't really come to grips with the problem of explaining how the ontologically troublesome events which seem to be additional to the neural events may after all be neural events. In implicitly denying the occurrence of private events the II-Identity Theory seemed to sweep under the carpet those events which pose the problem. And the I-Identity Theory seemed doomed because it tried to identify what may be private mental events with physical events.

Moreover, this division between the I- and II- Theories seemed to be inevitable, since it was not clear how we could recognise and sympathetically deal with the ontologically troublesome class of events without having to suppose they are private.

Thus, though Smart's original scientific Identity Theory clearly wasn't intended to be classed as one or other of the unsuccessful forms, there seemed to be no option but to conceive it as one or other of these.

But through the present account we can recognise the events which have been taken to be additional to physical events, while explaining how they are not after all private. (We will reassure ourselves of the latter presently.) And hopefully we can persuade the privacy objector to the Identity Theory that those events which he takes to be private and not plausibly identifiable with neural events are the events which we can explain are not private, and plausibly are neural events. In other words, there is hopefully sufficient common ground in his and our connotation of 'mental event' to make the Identity Theory a viable bridge by which he can cross over to monism.

The Identity Theory bolstered by the suggestions here in Part 4 I will call the augmented Identity Theory. We try to achieve Smart's original aim - to hypothesise that events which may seem to be additional to neural events, and for which we have not at present any intrinsic physical description though we can report their occurrence in a certain type of report, are in fact neural events - but we are able to avoid the fate of the I-Identity Theory: at best being forced to derive its only support for the perceptibility of mental events from the truth of the Identity Hypothesis.

Thus the augmented Identity Theory is more than just a II-Identity Theory, which fails to engage with the ontological problem, but without having to appear as a I-Identity Theory.

4.52 The main objections to the I-Identity Theory were summarised in 3.10. I will reconsider them in order.

Objection (i): This was a form of the incorrigibility objection. Using the family of predicates: I'_{ψ} for each ψ - where 'I'_{ψ} x' is read as:

'$\sim\Diamond$(A Asserts the proposition expressed by 'ψ' occurs as part of A's experience' and x does not occur)';

and 'Asserts' is read as:

'sincerely, and with linguistic competence, explicitly formulates and assents to' ;

it is supposed that, for each ϕ:

$$(x = \phi) \equiv I'_\psi x ,$$

$$\sim I'_\psi \phi ,$$

so $$\sim(\phi = \phi) .$$

To combat this objection I will essentially revive one of Smart's counters to Baier (cf. 3.51).

The incorrigibility objector has to and wants to allow that ϕ can occur as part of A's experience without A formulating or assenting to any proposition. But it seems to me that if it is true that for each ϕ, $I'_\psi \phi$, then for ϕ to occur as part of A's experience must <u>be</u> for A to formulate and assent to the relevant proposition. It is not clear how we can coherently allow ϕ to be an event which is <u>distinct</u> from A's assenting to a certain proposition, while maintaining that the occurrence of the former is logically rather than contingently connected to the occurrence of the latter. Essentially, the problems which we saw in Part 2 as in the end prevailing against the possibility of P-primacy are equally troublesome to the incorrigibility objection.

It is certainly not easy to conceive how one could assent to the proposition 'I'm in pain' (sincerely etc.), and not be in pain. It is significant I think that it is easier to conceive that one could assent to 'I was in pain then', when in fact one wasn't in pain then. This may be because what is really inconceivable is that one can be in pain and not be in pain. And what gives the mistaken element of inconceivability to the possibility that one could assent to 'I'm in pain <u>now</u>' and not be in pain, is only that one is trying to imagine <u>being</u> in pain and not being in pain - rather than trying to imagine having a certain thought and not being in pain. In the sense that one can <u>think</u> that one is in pain, or think that one was in pain, where the thought processes themselves don't "hurt", we must surely accept that it is <u>logically</u> possible that these thought processes occur even though "nothing <u>hurts</u>". Perhaps, because the thought process may accompany the pain, may occur at the same time,

it is not easy to separate these processes introspectively. But if they are separate processes, then in the sense that thinking one's in pain is not feeling the pain, we must allow that however contingently improbable, it is logically possible that the former occurs in the absence of the latter. (This sort of improbable and "inconsistent" combination of processes sometimes occurs in dreams, where in retrospect one realises that what one was thinking in the dream was curiously at odds with other aspects of the dream.)

By the same token, we must allow that the process of thinking that one was not in pain could occur even though there occurred a feeling of pain. We could not in this case easily say that while one was having the thought that one was not in pain, though a feeling of pain occurred, one had introspective evidence that one was in pain. But there could be evidence of various sorts that the feeling of pain occurred nevertheless - apart from neural evidence (i.e., if the "pain-event" was perceived), one could afterwards remember having felt the pain, and so on.

Clearly a lot more than I say here could be and needs to be said along these lines. At any rate, this is the sort of direction in which one would have to argue to combat the intuitive feeling that there is a "subject" in a logically privileged position with respect to mental events it is "aware of", and that there is some sort of"unity" embracing the occurrence of mental events which occur at a certain time. (I will mention this issue again in Part 5.) It would be envisaged on the Identity Theory that in principle we could by neural tampering induce the odd combinations of events which are required to be logically possible on this view.

Objection (ii): This was essentially a form of the imperceptibility objection. Through the J-linguistic treatment we can explain the intuitive factors underlying the contention that mental events are in principle imperceptible, while showing that such a contention cannot coherently be justified. (Cf.4.42, especially (5c).) Thus we can deny the first premiss of Objection (ii), the premiss that $(x = \mu) \supset \mathrm{IMP}x$, by denying that $\mathrm{IMP}\mu$.

Objection (iii) : This was a form of the location objection. Clearly mental/neural events cannot be put into detailed spatial relations with other physical events until we can perceive them. But once we have overcome the crucial imperceptibility objection, the location objection falls with it. And we can get a heuristic idea of the sort of confusion through which it may be thought that mental events cannot be located.

Suppose a wff in a J-language L_K consists of strings of numerals. And suppose that in, say, k^i : '3040', the arrangement of the numerals maps an arrangement - we may call it "spatial" - in the J-desiignatum. The analogous confusion is to try in vain to compare, say, '3040' with its own third numeral, and assume as a result that k^i was not susceptible to the "spatial" relations which may obtain in the J-desiignatum. But if k^i is J-desiignated in some wff, that would reveal how it is "spatially" oriented.

So again we deny the first premiss of Objection (iii).(We deny that S'$\not\phi$.)

Objection (iv): This objection demanded an explanation: essentially to explain what seems to be additional to ϕ is not in fact additional to ϕ. Hopefully we now have at least an indication of how such an explanation might go.

We may reassure ourselves here of the empirical nature of the hypothesis that $\not\phi = \phi$.

We have to regard as at best of very dubious logical coherence the view that mental events could be in principle imperceptible. For if they are in principle imperceptible, they are either like imperceptible pink elephants: no one could perceive them or have any evidence for their existence; or "someone" has privileged evidence for their occurrence. But the privileged position is not then a contingently privileged position - and so we cannot give any clear sense to the independent existence of mental events. We could not give a coherent account of the distinction between thinking one's in pain and being in pain. Yet this sort of distinction is required if there can be any "structure" to "consciousness". This is of course in essence a form of the Private Language Argument.

So we have the case where mental events for the occurrence of which we may have introspective evidence (though not logically privileged eviedence), must in principle be imperceptible. And now the empirical content of the augmented Idented Theory is essentially in determining what is to count as perceiving mental events - in other words, what are to count as physical descriptions of such events. The likely candidates seem to be neural descriptions.

We must note that although the augmented Identity Theory involves an empirical hypothesis, it involves a very unusual one. For normally empirical identity hypotheses take the form of supposing that something with a certain physical description has also the other physical description given in the hypothesis. But on the augmented Identity Theory, we hypothesise that events which occur in us for which we don't at present have intrinsic physical descriptions, have neural descriptions as their intrinsic physical descriptions.

Objection (v): This objection involved the notion of Residual Privacy. Now, we have seen that on the present account mental events are not in principle private in the sense of being in principle imperceptible to others and having a "subject" in a logically privileged position with respect to them. And we have seen that the relevant notion of incorrigibility must and can be abandoned. The only symptom of "privacy" we have left is Residual Privacy.

Now, the notion of Residual Privacy clearly is intimately related to the notion of J-privacy: in particular, to the notion of a domain W not being univocally accessible to J_K. The notion of Residual Privacy essentially picks on the availability of an inverting function for visual imagings. And it is thus essentially a manifestation of J-privacy for the particular structure J_A. Residual Privacy, like the more general J-privacy, is compatible with and may be expected to apply to a monistic domain: the notion of J-privacy is not in any way peculiar to "mental" events.

We may assume that V is not totally accessible to J_A: this would be making an assumption about the j-function which J_A operates. But we know, through the availability of the relevant inverting function, that V is not univocally accessible to J_A. And thus we know that a wff a^i cannot capture

all the information there is to be captured regarding $j(a^i)$. But all this is to be expected if visual imagings are neural events.

But what we must stress is that the notion of J-privacy does not require the possibility of the P-primacy of Part 2. The distinction between two isomorphic visual imagings is not in principle private to one person: for if the visual imaging can be perceived, someone else may be in as good a position to distinguish between isomorphic visual imagings as the person they occur in. It was only when visual imagings were supposedly additional to any perceptible neural events, that we could construct an example where, say, from birth a person's visual imagings had all been lateral inversions of what in fact they have turned out to be - but with no perceptible manifestation; for now we have to suppose that there would be a perceptible manifestation: through the neural structure.

So although we have the notion of J-privacy in the augmented Identity Theory, a notion which retains the essential features of Residual Privacy, this does not land us with the problematic P-primacy.

4.53 The restrictions which Wittgenstein's account imposes on the notion of mental events leaves these in an uneasy limbo. For on the one hand, it is not possible on Wittgenstein's account to allow that there are mental events which are in principle imperceptible. But on the other hand, there is considerable pressure for supposing that we can be introspectively aware of mental events occurring in us, and that these are in principle imperceptible.

But through the augmented Identity Theory we can hopefully resolve this uneasiness, by exhibiting the mental events of which we can be introspectively aware as only contingently imperceptible. Moreover we retain the feature of Smart's view, that these events as a matter of fact have neural descriptions; but in contrast to Smart's view, on the augmented Identity Theory mental events are physical as a matter of logic.

I will now in Part 5 discuss some features of the augmented Identity Theory in a more general context.

NOTES TO PART 4

[1] The class of imagings does not of course exhaust the class of private mental events thus conceived. But it is through a confusion involving perceptual events that we shall expose the dualist's errors.

[2] Wittgenstein was describing contexts in which a sense could be given to talking about the locality of thoughts - but this setting for the passage cited need not concern us here.

[3] Both 4.3 and 4.4 are rather more compressed than I would like - though I hope not to the point of unintelligibility. In the context of this thesis I can only attempt to indicate the direction which a more comprehensive account would follow.

[4] The definition here is based on that given in Rotman and Kneebone, pp.31-2.

[5] I will later draw the connection between the notion of Residual Privacy (cf. 2.3)and J-privacy.

[6] If L_X is a graphic J-language for U, then we could say that X is a magnifying domain for U - by analogy with "amplification" by means of, say, a magnifying glass as opposed to, say, "amplification" through a digital print-out. But we need not go into this here.

[7] We shall see in Part 5 that these twin confusions have been compounded in a rather interesting way in the views of Clifford and Prince - leading to the identification of the imperceptible consciousness of a person with an imperceptible "thing-in-itself". In one sense the compounding of these two wrongs makes a right: in that the result is a monistic view.

[8] To prove that $(K' \neq W') \supset W$ is J-private from J_K in L_K, it is sufficient, by Theorem 2, to show that for every W there is an inverting function. The latter may very plausibly be the case, at least within certain domains.

[9] We have to make the assumption that J_A scans A' through the amplifying structure J_X, since contingently we cannot perceive neural events directly. But this assumption is not important for my present purpose.

[10] Unless on a non-null interpretation $a(a_n)$ J-designates itself - but we will prove that such a case cannot arise.

[11] Though the occurrence of pains and kinaesthetic sensations may most simply be regarded as having events inside the body as their J-designata.

[12] We may plausibly suppose that in lower animals this capacity for higher cerebration is absent.

[13] The dualist may come back at us and dispute the possibility that a purely physical structure could be introspectively aware of events occurring in it in the way that he can be introspectively aware of his mental events. But I'm not sure what sound basis this objection could have.

[14] Though we will not formally extend our result to J-languages generally, it will be fairly clear that the result will apply generally to J-languages in which non-trivial self-reference is possible.

[15] The corner marks are used for quasi-quotation (cf. Quine (1), pp. 35-6): $\ulcorner a(\underline{v}^i) \urcorner$ is the expression formed when $\underline{v}^i$ is put in place of "$\underline{v}^i$" in "$a(\underline{v}^i)$". I.e., the constant contextual background "a( )" is quoted, and the unspecified expression $\underline{v}^i$ is imagined written within the brackets. In general, all the constants inside the corner marks are quoted, but all the variables are imagined as replaced.

[16] Both this operation and the corresponding operation of "N" could be recursively effected by J_X.

[17] In case it is wondered why we don't have $x_e^{G'}$ as the J-designatum of $a_e^{G'}$, it should be borne in mind that in principle when J_A is scanning X' we construe X' as part of V', and though $j(a_n) = v_n$, not necessarily $j(a_n) = x_n$. Here although X' and V' coincide for all intents and purposes, we nominally retain the distinction.

Note also that we say that $a(x^i)$ J-designates x^i in L_A, <u>not</u> in L_{A_1}, despite the fact that ${}^{A_1 \to X}J(a(x^i), x^i)$, since we construe x^i as a member of V, ${}^{A_1 \to X}J$ being a restriction of ${}^{A \to V}J$, not giving - in the sense of not being a <u>resultant</u> relation - a new interpretation.

[18] More generally we may note the resemblance to a self-referential aspect of Cantor's so-called "diagonal argument". This argument appears in its most general form in the proof of Cantor's Theorem, which states that the power

set of a set S (the set of all subsets of S) has a greater cardinal number than the set S - in effect, that there cannot be a bijection from S to its power set. Cantor's Theorem is proven by considering the respectably specified set $P = \{x \mid x \in S \text{ and } x \notin f(x)\}$, where f is a function from S to its power set. (Thus f(x) is the subset of S "correlated" with x through the function or mapping f.) P is a subset of S; but when we consider the possibility that f(x) is P itself - this is where the self-referential aspect comes in - we find that whatever f is like, $P \neq f(x)$, since $P = f(x)$ implies that $x \in P$ iff $x \notin P$. So P must be "left out" by the mapping f.

[19] But if we have a suitable "second order" function we may escape this restriction if the "second order" function does not prescribe one and only one designator for each expression of S. This is not sufficient to permit self-designation, but necessary for this. In the case of d_2 , $d_2(E)$ designates an expression F of S only if F is of the form $\ulcorner Ed_1(E) \urcorner$.

[20] We could however envisage J_ζ strengthened with the diagonal function facility, and thus envisage "second order" self-depiction.

[21] Cf. Turing, p.16.

[22] Cf., for example, Rogers, p.136.

[23] This is obviously a rather contrived interpretation: it must be remembered that where we are particularly interested in the arithmetical interpretation of the outputs of a Turing machine, it is of particular interest to transpose the blindspot into an arithmetical limitation. In our case we originally were not at all interested in an arithmetical interpretation. But we must expect the arithmetical interpretation to be available, since it is part of what is meant by saying that a structure operates recursively that in principle it could be arithmetised.

[24] In principle this need not be an arithmetical limitation for J_A^+ : since J_A^+ might have other ways of computing the function g than by J_A's outputs in L_A . The limitation is only on J_A interpreted as performing arithmetical computations.

PART 5

THE AUGMENTED IDENTITY THEORY

I will make some general remarks about the augmented Identity Theory in 5.2, and mention some problems for it in 5.3. But first I will discuss a related strategy which has been used to support ontological monism - a strategy related to the augmented Identity Theory in its use of a realist distinction.

5.1 The Right Distinction Wrongly Drawn

> There are two parties to every observation - the observed and the observer. ((Eddington, p. 30.))

The crucial distinction between the observer and the observed has been used previously in support of ontological monism. The fact that it has been so neglected in this context makes me rather pessimistic about the chances of the present account contributing any more persuasively.

But there are reasons perhaps why the previous proposals incorporating this distinction have failed to graft successfully into philosophical discussion. First, these proposals tend to occur as part of views - notably the "mind-stuff" theories of Clifford and Prince - which appear at least superficially to be diametrically opposed to materialism. And secondly, in most of the writings I know of where it is possible to recognise elements of the view I am advocating, the expression of these elements is couched in terms containing a very familiar mistake about meaning - involving roughly a pre-Stage I view. Both these aspects can be illustrated in connection with Clifford's view.

5.11 .. the inferred existence of your feelings, of objective groupings among them similar to those among my feelings, and of a subjective order in many respects analogous to my own, - these inferred existences are in the very act of inference thrown out of my consciousness, recognised as outside of it, as not being part of me. I propose, accordingly, to call these inferred existences ejects, things thrown out of my consciousness, to distinguish them from objects, things presented in my consciousness, phenomena. ((Clifford (2), p. 54.))

Clifford takes "ejects" to include not only "groupings" in other people's consciousness (as in the citation), but also, as "non-personal ejects", things-in-themselves. And he takes a "material object" to be what we would more comfortably describe as a mental-perceptual image of a material object:

> This room, the table, the chairs, your bodies, are all objects in my consciousness; ((Ibid., p. 56.))

Clifford is making a simple mistake with regard to the meaning of "this room", etc., partly to clarify his views, but partly no doubt as a hangover of the same tendency in Locke and Kant. We just don't see Clifford's table - as construed by Clifford the table would be a private object, and "table" would be a term in a private language. We would be much more comfortable saying that we perceived Clifford's table-eject.[2]

The "stuff" of which consciousness is a complex, Clifford calls "mind-stuff". And Clifford tries to show that his things-in-themselves, "external reality" must also be composed of mind-stuff. The interesting thing about Clifford's account of this is that we have a demonstration that there is no need to posit two sorts of things - Clifford ends up with a monistic conception, except that because of his mistake about meaning he finds sense to maintaining that the world is made of mind-stuff, and not matter.

Clifford considers the case where he is observing a man looking at a candlestick, where he can observe the cerebral processes going on in the man. As we shall see, Clifford's confusion of meaning really is

confusing here - it may be helpful to transpose Clifford's terms to ones we might find less confusing. For Clifford's 'cerebral image', try '(Clifford's) visual image of the man's neural configuration of (relevant neural state when looking at) the candlestick'. For 'candlestick' try '(Clifford's) visual image of the candlestick'; for 'external reality' try 'candlestick'; and for 'mental image' try 'the man's visual image of the candlestick':

> Suppose that I see a man looking at a candlestick. Both of these are objects, or phenomena, in my mind. An image of the candlestick, in the optical sense, is formed upon his retina and nerve messages go from all parts of this to form what we may call a cerebral image ... This cerebral image ... is a material or physical fact, therefore a group of my possible sensations, just as the candlestick is. The cerebral image is an imperfect representation of the candlestick, corresponding to it point for point in a certain way. Both the candlestick and the cerebral image are matter; but one material complex represents the other ...
>
> Now the candlestick is not the external reality whose existence is represented in the man's mind; for the candlestick is a mere perception in my mind. Nor is the cerebral image the man's perception of the candlestick; for the cerebral image is merely an idea of a possible perception in my mind. But there is a perception in the man's mind, which we may call the mental image; and this corresponds to some external reality. The external reality bears the same relation to the mental image that the (phenomenal) candlestick bears to the cerebral image. Now the candlestick and the cerebral image are both matter; they are made of the same stuff. Therefore the external reality is made of the same stuff as the man's perception or mental image, that is, it is made of mind-stuff. ((Clifford (2), pp. 70 - 71.))

Thus Clifford implies but does not quite give the more striking relationship (in Clifford's terms): the cerebral image stands to the mental image as the candlestick stands to the external reality. Or transposed: Clifford's visual image of the man's neural configuration stands to the man's visual image as Clifford's visual image of the candlestick stands to the candlestick.

But of course, Clifford's conclusion in terms of mind-stuff does not coherently follow in our transposition. (And as I will mention later, nothing accurate is gained by suggesting that the augmented Identity Theory advocates panpsychism.[3])

As Clifford notes, a hypothesis resembling his mind-stuff hypothesis is mentioned by Kant - for instance:

> I may further assume that the substance which in relation to our outer sense possesses extension is in itself the possessor of thoughts, and that these thoughts can by means of its own inner sense be consciously represented. In this way, what in one relation is entitled corporeal would in another relation be at the same time a thinking being, whose thoughts we cannot intuit, though we can indeed intuit their signs in the ((field of)) appearance. Accordingly, ... the very same being which, as outer appearance, is extended, is (in itself) internally a subject, and is not composite, but is simple and thinks.
>
> But, without committing ourselves in regard to such hypotheses, we can make this general remark. If I understand by soul a thinking being in itself, the question whether or not it is the same in kind as matter - matter not being a thing in itself, but merely a species of representations in us - is by its very terms illegitimate. ((Kant, p. 340 (Second Paralogism).))

For Kant, though, there is an uneasy and supposedly impermeable conceptual barrier involved in his distinction between phenomena and noumena. Thus a hypothesis of Clifford's sort would have no coherent rendering in Kant's view.[4] And indeed, elsewhere Kant suggests that a conceptual misapplication would be involved in the supposition that there is a basis for maintaining either transcendental dualism or transcendental monism (cf. Kant, pp. 351 - 352 (Fourth Paralogism)).

5.12 Prince's view is essentially the same as Clifford's. It contains roughly the same difficulty about meaning: 'The only thing we know is our sensations.' (Prince, p. 192). And:

> The object, then, does not exist as outside of us, but is only a bundle of our sensations. Undoubtedly something exists outside of us which is the cause of these sensations in us. This something has been called the thing-in-itself, but its nature is unknown to us. ((Ibid., p. 192.))

On the other hand Prince exhibits rather more clearly than Clifford the sort of device by which we may collapse dualism, though his monism, like Clifford's, contains the one substance - mind:

> ... instead of there being one substance with <u>two properties</u> or 'aspects' - mind and motion - <u>there is one substance, mind</u>; and the other <u>apparent</u> property, motion, is only the way in which this

> real substance, mind, is apprehended by a second organism: only the sensations of, or effect upon, the second organism, when acted upon (ideally) by the real substance, mind. ((Ibid., p. 191.))

We have here the observer/observed distinction but again infelicitously expressed. Because, quite apart from anything else, if we take Prince's (and Clifford's) style seriously, he cannot express the distinction he wants by means of it: the "second organism" will be a "group of his sensations" - in order to make his distinction work he must talk about what he regards as "things-in-themselves" in the way in which he allows himself only to talk about his sensations. Nevertheless, if we forgive all this and reinterpret him, we can recognise a number of valuable points; in particular, the sort of distinction I describe as contrasting a J-designating and a J-designated event, and the notion I describe as subjective dualism:

> ... how would this sensation of pain appear to us if we could apprehend it through our senses, and through the sense of sight in particular ...? The answer is, Only as all other activities in matter appear to us, namely, as motions, undulation, etc. If, then, these hypothetical conditions were the facts, it would be easy to understand how mental states can become 'transformed' into physical disturbances, and vice versa, because there is no transformation about it. There would be in this case only one thing, mental states, which would appear as physical activities when viewed (ideally) through the senses, as tremors if viewed through sight.
>
> In other words, a mental state and those physical changes which are known in the objective world as neural undulations are one and the same thing, BUT THE FORMER IS THE ACTUALITY, THE LATTER A MODE BY WHICH IT IS PRESENTED TO THE CONSCIOUSNESS OF A SECOND PERSON ... ((Ibid., pp. 193 - 194.))
>
> Here, then, lies the parallelism of the phenomena: your consciousness or pain is the correlate of my apprehension of this consciousness as neural vibration. The parallelism is between your consciousness and my consciousness of your consciousness, or, what is the same thing, between the consciousness in you and the picture in my mind of neural vibrations. The former is the reality, the latter the symbol of it. ((Ibid., p. 195.))

(In the first of the two citations above Prince is presumably using 'matter' to mean the same as his 'actual matter' (cf. Prince, p. 192), roughly, to refer to things-in-themselves.)

Although we can forgive the mistakes about meaning involved, and hence find it unnecessary to conclude that the things in the world are all mental rather than physical, it should also be emphasised that to suppose that the events occurring in the world are <u>really like</u> the events going on in me in any other sense than that neural events are "like" any other physical event, is incoherent: the only coherent sense in which physical events are "like" my mental events is the sense in which there is no difference between my mental events and other events, other than in respect of what we understand by "physical difference".

Russell expresses the crucial distinction, but he again is dogged by an infelicity of meaning:

> Then, again, there is the argument about brain and mind. When a physiologist examines a brain, he does not see thoughts, therefore the brain is one thing and the mind which thinks is another. The fallacy in this argument consists in supposing that a man can see matter ... His percept when he looks at a brain is an event in his own mind, and has only a causal connection with the brain that he fancies he is seeing. ((Russell, pp. 244 - 245.))

Schlick's rendering of this sort of point is perhaps the most accurate I have come across:

> The gravest error which can be made in interpreting the psycho-physical problem, and which, incomprehensibly, is repeatedly being made, is the fact that, unnoticed, the perceptions or images of the brain processes are put in the place of the brain processes with which the mental processes are supposed to be identical. The perceptions themselves are reality experienced, are themselves mental processes, but they belong to another individual, that is to the man who contemplates the brain of the first individual, and they are naturally in no way identical with the experiences of the first individual; they are not 'parallel' with them, but rather are causally dependent upon them, for what I perceive in a man's brain will depend, on principle, upon what is happening in his consciousness. ((Schlick, p. 306.)) 5

<u>5.13</u> We see then through the views of Clifford and Prince that even if the realist distinction is drawn incorrectly in a manner which would seem favourable to the dualist, a case can still be made out for monism.

For although Clifford and Prince make the distinction incorrectly, at least they are very much aware of it. Whereas, as I have suggested, dualist arguments tend to require a realist distinction implicitly, while at the same time overlooking it at the crucial stages. But of course in the augmented Identity Theory an attempt is made to discard the various problematic notions which may attend a realist distinction - in particular, the notion of a subject who is aware of his mental events as they occur; and the almost inevitable concomitant of this: the unperceived or "indirectly" perceived thing-in-itself.

We should note also that until Relativity Theory the realist distinction was largely overlooked in physics. Of course, the notion of an "observer" in Relativity Theory does not require a "psychological" connotation - the observer may be "just" a physical structure which "measures" certain properties of the observed physical event in that it is causally affected by that event. But if we overlook the distinction between the events in the observer and the events "measured" or observed, we may easily confuse what really are observational parameters, such as spatial position and temporal position, with "objective" relations between the events observed: relations about which all observers would agree, and which are <u>functions</u> of the observational parameters.[6]

As I have mentioned already in Part 4, it is easy to overlook the observer/observed distinction. It is a distinction which we may make after we have already learnt to respond to our environment - indeed, being aware of it may be obstructive to our normal response. But it is a distinction which we may be reminded of when we consider:

(a) The perception of a very distant event, for example, happening on Sirius - an event ordinarily describable as taking an appreciable amount of time to cause the perceiving event. It is easy though to fall into the trap of supposing that if we are looking, at time t, at Sirius,

and we see Sirius explode, we <u>see</u> a spot-of-light-expanding-at-time-t; we are looking and seeing at time t, but <u>what we see</u> is Sirius exploding at t - 9 years; and the J-designations of the Sirian event occur at t. We don't see the spot-of-light-expanding-at-t any more than we see a visual imaging J-designating a nearby event. But we can say that we see, at t, Sirius-exploding-at-(t - 9); and in any sense in which we say that we <u>see</u> the spot of light, the spot of light <u>is</u> Sirius.

(b) The perception of an event ordinarily describable as involving the movement of something at a speed which is appreciable compared to the fastest speed at which a cause can manifest itself through an effect - empirically, the speed of light. For in this case our physical description of such an event would be at odds with Newtonian mechanics, in which the observer/observed distinction is not taken closely into account.

(c) The perception of an event which we have reason to suppose is appreciably similar to the perceiving event, or other events occurring in the perceiving structure.

I merely hint here at a complex engagement between the augmented Identity Theory and Relativity Theory. Clearly, mental/neural events, like all physical events, are subject to relativistic considerations. We should note in particular a simplification present in the J-linguistic treatment of visual imagings, which is exposed by relativistic considerations: we must assume that the ordering relation on what I described as a J-linguistic sequence of visual imagings (/neural events) is not given by a <u>temporal</u> relation, but by a relativistic interval-relation. For the <u>times</u> at which neural events occur, and simultaneity relations on neural events, may be expected from Relativity Theory to vary depending on the frame of reference of the observer of the neural events. Thus temporal and simultaneity assignments to a neural structure by an observer moving at a very high speed relative to the neural structure would differ

appreciably from the assignments of an observer who is stationary relative to the observed neural structure. So the ordering of visual imagings in their J-linguistic sequence, which we must take to be invariant, is not a temporal ordering, since this latter may vary depending on the frame of reference of the observer perceiving the visual imagings/neural events.

This suggests that not only can we not assign a spatial position to mental events until we perceive them - we also cannot assign temporal positions to them until we perceive them.

But what about the feeling that we can at present specify mental events as occurring at temporal positions? Well, what we do at present, I suggest, is (quite reasonably) assign to mental events (which contingently we can't perceive) the same temporal positions we assign to nearby events - events occurring near or happening to our bodies - implicitly assuming that mental events are occurring in our bodies (rather than several light-years from our bodies, for instance). In other words, we are implicitly assuming that if we could perceive our mental events they would be seen to occur in our bodies.[7]

I will not go into this complex issue any further here, though it will come up again briefly later on.[8]

5.2 The Augmented Identity Theory and Physicalism

On the augmented Identity Theory (a) we are purely physical structures; and (b) what are described as mental events occurring in human beings are physical events occurring in their brains or central nervous systems. I will first stress some general consequences of (a); and then I will discuss some aspects of (b) regarding the form of the Identity Hypothesis.

5.21 The events which constitute what may seem to a person to be his

"private stream of consciousness" are physical events on the augmented Identity Theory, and therefore differ from events we uncontroversially classify as physical only in respect of physical complexity and structure.

Does this mean then that physical events are "like" our mental events - trains of thought, feelings of pain, and so on - and that we are led to a form of panpsychism? Well, it would be quite misleading to describe the present view as a form of panpsychism.

In the first place, a panpsychist view may be taken to suggest that the real constituents of the world are mental, not physical. This is clearly not the case on the present view: the real and only constituents of the world are physical. There are just more or less complex and more or less differing physical events. Moreover, although we classify certain physical events occurring in human beings as mental events, we cannot suppose on the present view that there is a well-defined class of physical events: those which are mental events. For if we make the plausible assumption that there is a continuity in the possible degrees of complexity of physical structures, we cannot suppose that there is a sharp dividing line distinguishing physical events which can be regarded as mental from those which cannot. This is not altogether explicit in Smart's view, though he does suggest that on his approach to materialism we don't have to ask the troublesome question: 'How could a non-physical property or entity suddenly arise in the course of animal evolution?' (Smart (4), p. 168).

Secondly, it might be thought that a way to panpsychism is: 'Whereas before I thought physical events were radically different from my mental events, now I understand that they are not - so they are like my mental events after all'. But a physical event <u>may</u> be like the mental events occurring in us <u>only</u> in that it's <u>physical</u>. Furthermore, it may be as different as it is different (physically). A physical event

occurring in a vastly complex neural structure is as different from an event occurring in a single neurone as the neural structure is different from the structure of a single neurone. And both these events will be different again from an event occurring in a bit of dust, or an atom. Nevertheless, to "produce" an event such as a train of thought, or a feeling of pain, all we have to do is combine atoms to form neurones, neurones to form a cerebral structure, and then an event occurring in the cerebral structure may be a train of thought or a feeling of pain. But to suppose it would be saying something further and informative if we said that an event in a single neurone is "like" a mental event would be like supposing that having said that a diamond is made only of carbon atoms, we may usefully say then that a carbon atom is like a diamond.

5.22 I mentioned just now in 5.21 that we cannot expect there to be a sharp dividing line distinguishing a special class of physical events: those which are mental events. By the same token, we cannot expect there to be a sharp dividing line distinguishing structures which we would judge to have minds from those we would judge to be mindless. That's to say, we must accept that there doesn't have to be a yes/no answer to every question of the form 'Does X have a mind? '. We can make a point applied to "mind" similar to Parfit's point applying to personal identity (cf. Parfit, p. 142), that we could in principle describe all there is to describe about X without having to, or necessarily being able to, answer the question 'Does X have a mind?'.

But we can say this: on a monistic view in which there is not an ontologically distinct class of mental events additional to physical events, there is in principle no obstacle to the view that a non-neural physical structure - an electronic structure, for example - could be said to have a mind in the same sense we are said to have minds. And in

principle there would be no obstacle to the view that mental events could occur in such a structure.

For the main obstacle to identifying a feeling of pain or a train of thought with a neural event was not that the latter was neural, but that it was physical. But now if there is no obstacle to identifying a train of thought, say, with a neural event, we may also be happy to envisage describing some electronic event as a train of thought.

Certainly, we have to say that there occur in some physical structure A, whether neural or electronic, events physically resembling a human being B's mental events in proportion to the degree to which the former resemble B's neural events. But if we are unable to examine closely enough the detailed physical events involved, we have to and can compare the events occurring in A and B at least to some extent by comparing A and B's "gross" behaviour.[9] This may hide certain differences in their physical mechanisms - but such detailed physical features are largely irrelevant to the basis on which we give meaning to mental attributions, and on which we compare mental events in different human beings.

As was discussed in Part 2, the way we learn the use of expressions describing and referring to mental events is such that the intrinsic physical properties of these events could not enter into the meaning of those expressions - for the simple reason that contingently we don't perceive these events. In other words, it is not part of the meaning of 'train of thought', 'visual imaging', 'feeling of pain', that there is a neural description for trains of thought, etc.. When 'train of thought', say, is used in a reference to a certain neural event, this involves an indirect extrinsic specification based on gross behavioural contexts; and the intrinsic nature of the sorts of event which qualify as trains of thought is not important in determining the use of 'train of thought'.

We succeed in referring to contingently imperceptible mental/neural events in an indirect manner, rather like the way Smith's murderer may be referred to by the expression just used without demanding detailed intrinsic qualifications on the part of Smith's murderer.

Most ordinary mental concepts get their significance from gross behavioural considerations, rather than from considerations of physical detail. And such concepts could thus sensibly apply to comparably behaving structures regardless of their detailed mechanisms. We may say analogously of a motor vehicle that it accelerates, turns a corner, brakes, regardless of whether it is powered by an internal combustion engine, an electric motor, or jet propulsion; in contrast, 'it revved up' may be applicable to a vehicle powered by the first two mechanisms, but not the third. And similarly, a neural description of a train of thought would only be applicable to a neural structure.

Although we may specify a vehicular "inner process" by 'it suddenly lost power', we may also have available a detailed physical description of this inner process. But contingently we have not at present any detailed physical descriptions of mental events specified through gross behavioural concepts.

I will again illustrate these points by means of an example involving a chess machine. (I have used this sort of example in 3.9.)

Suppose I am playing chess against an electronic chess machine, and contingently I cannot investigate, describe, explain or predict the behaviour of the machine in terms of its electronic micro-states. Nevertheless, in spite of the fact that I cannot observe the causal antecedents to each output, I can still describe, explain and predict (or attempt to predict) the processes occurring in the machine in terms of the gross behavioural context, and the gross regularities and tendencies displayed through the machine's outputs. I can give an

account of the machine's part in the game in terms of tags which when unravelled are intelligible in terms of what is readily observable and extrinsic to the electronic mechanisms.

Thus I may describe an output as 'It moved its king pawn'. For this output I will have available a physical description - e.g., of certain signs appearing on a visual display. The relation between the two sorts of description here is very much like the relation between a physical description of a wedding and its description as a wedding.

I may offer as an explanation for one of its moves: 'It decided that I was going to exchange bishops, so it moved its bishop to avoid an unfavourable exchange'. Because of the comparative infrequency of outputs compared to the complexity of the electronic processes going on in the machine, I may be unable to date the machine's decision more precisely than at some time between my last input and its last output. But an examination of the electronic events might enable me to pin-point more precisely when the decision occurred. As regards the form of the explanation given - it is clearly in terms of goals which I may have observed the machine trying to achieve. (Of course, if I know already that the machine is a chess machine trying to win games of chess, I can guess some of these goals. In this respect this example differs from a case where I have to determine the goals empirically from the gross behavioural considerations.)

All this is of course quite compatible with there being a detailed description for the causal antecedents of every event occurring in the machine.

The relation between an explanation in terms of electronic causal antecedents and one in terms of goals in chess is naturally complex and indirect: there is no clear generalisable engagement between the two (as was mentioned in the discussion of Davidson's view in 3.9).

Clearly the second sort of explanation is indispensable to any immediately useful account of what the machine did in the chess game. And we can give a description and explanation of this without any knowledge of the detailed physical mechanisms involved in the working of the machine - without even knowing that they are electronic.

Conversely, a scientist could in principle describe and explain the events occurring in the machine through intrinsic physical descriptions: without knowing anything about the game of chess. But the fact that the machine can be interpreted as playing a game of chess, and that therefore it could be described against that background, hardly presents a scientific problem for the scientist: we could hardly say that because the scientist didn't know anything about chess, and therefore could "only" explain every single physical event in the machine in terms of physically described causal antecedents, that the machine still presented him with a scientific problem. Because if it is supposed that there is a scientific problem thereby presented, then the scientist would have to prepare himself for every possible background against which to describe any single physical event, before he can be said to have achieved any scientific completeness in his investigation and explanation of that single event.

And even if the scientist is thereby doomed to scientific incompleteness, we can at least say this: although a structure which behaves in a very complex manner is more liable to be described against specialised backgrounds, we can see that neither purely physical human beings nor any other comparably behaving structure, every event in which could be explained in physicalist terms, present a special problem to the scientist - a problem other than simply in the degree of difficulty of physical investigation - which thereby distinguishes them from other physical structures by presenting a unique logical resistance to scientific

accommodation. For any sense of scientific incompleteness which could be maintained as remaining is clearly applicable to the description and explanation of any physical event whatsoever.

I will say a little more regarding this last issue presently. For the moment we may note that in the case, for example, of an appropriately behaving electronic structure, we could specify events occurring in it as trains of thought, sensory imagings, and so on - only in this case these events would be electronic, not neural. And whether or not we know enough about the physical mechanisms in such a structure to describe and explain its behaviour in terms of causal antecedents and so on, we could still describe, explain and attempt to predict its behaviour as we do at present for our own behaviour.

But the really striking difficulty in accepting that an electronic structure might be "conscious" disappears - this difficulty is after all the same difficulty which faces the purely physical neural structure: 'I look at a brain - neural or electronic - and all I see is grey matter or electronic circuits; nowhere do I find the stream of consciousness. I have to accept then that the latter stream must be additional to the neural events. But I cannot easily accept that there is a corresponding stream additional to the electronic events.' But we have exposed the sort of confusion which creates this difficulty: the confusion involves looking for neural or electronic events amidst one's J-designations of them. The stream of consciousness is just the physical stream, whether neural or electronic.

5.23 We must thus accept for the augmented Identity Theory the categorial difference between most mental and physical discourse. Thus we must be careful to make sure that what we specify as mental which <u>is</u> physical on the augmented Identity Theory, involves a sensible categorial "matching": this is achieved by identifying mental <u>events</u> with physical <u>events</u>.

(We could have mismatching, in identity hypotheses generally, without supposing there are categorial differences in the "discourse" couching the descriptions figuring in the hypotheses: if for example we tried to identify objects with events, and so on. The point is that whatever the categorial differences between mental and physical discourse, references to events can be made through both, and this allows the engagement of these two forms of discourse in identity hypotheses.)[10]

Clearly, on the augmented Identity Theory, particular dated mental events are particular dated physical events; and in human beings particular mental events are hypothesised to be particular <u>neural</u> events.

But we must be careful when considering hypotheses of the form:

'A mental event of type M is a neural event of type N.'

For we must not interpret the augmented Identity Theory as suggesting that we could obtain bi-conditionals of the form:

'A mental event of type M occurs iff a neural event of type N occurs.'

To expect such <u>general</u> bi-conditionals of this form would be as absurd as expecting that, for example:

'(A physical object) x is a chess knight iff x has (physical property)P.'

For the qualifications required of a physical object to qualify as a chess knight are such that they just cannot be rendered in this manner. And for an event to be a feeling of pain, or a visual imaging, it has to satisfy certain gross behavioural restrictions, which we cannot expect to allow bi-conditionals of the form:

'A visual imaging J-designating v^{ι} occurs iff a neural event of type N occurs.'

'A feeling of pain of the sort ... occurs iff a neural event of type N occurs.'

We cannot envisage bi-conditionals of this sort, because apart from anything else, if an electronic structure, say, behaved as we do

(and had visual organs and so on), a visual imaging in the electronic structure would be an electronic event.

On the other hand, just as we might expect that:

'(A physical object) x is a chess knight on a particular type of board iff x has (physical property) P.'

we may expect that:

'A visual imaging J-designating v^i occurs in a human being iff a neural event of type N occurs in the human being.'

(However, we cannot expect <u>every</u> specification of a type of mental event to yield bi-conditionals for human beings: many such specifications would be too vague or idiosyncratic: for example, 'a sudden decision to invest one's money', 'a visual imaging J-designating a good French actress buying a second-hand car'.)

So, on the augmented Identity Theory (a) every particular mental event occurring in a human being is a particular neural event occurring in him. And (b) if we can genuinely specify with sufficient precision a type of mental event M, then we should expect in principle to obtain bi-conditionals of the form:

'ψ occurs of type M in a human being iff ϕ occurs of type N in him (and $\psi = \phi$).'

Now, many if not most mental ascriptions do not assert the occurrence of mental events; and most statements asserting the occurrence of a mental event cannot be analysed in a form 'an event of type M occurred' which could yield the sort of bi-conditional just mooted.

And there is a tendency to regard the absence of law-like relationships, meaning reducibility and numerical correspondence between the concepts in mental discourse and the concepts envisaged in physicalist discourse as a crucial problem for the scientific accommodation of mind.

But as I have already suggested, this really cannot be construed

as the crucial problem for the scientist. The crucial possibility allowing in principle the scientific accommodation of mind is the possibility of a description and explanation of all the events occurring in human beings and in which they partake, in physicalist terms. Thus it is ontological dualism, not conceptual dualism which presents the problem for the physicalist.

The scientist can accommodate with little difficulty the fact that gross behavioural concepts are superimposed on a substratum susceptible to physicalist description, without requiring this superimposition to be attended by any strict relations of meaning, or any law-like relations between the two sorts of concept.

I have already hinted at some justifications for this view. But I will try now to consider in a little more detail an objection to the view that the conceptual dualism is no problem for the scientist.

Suppose that a scientist has his homogeneous system of description and explanation for every event occurring in a person and in which the person partakes - in physicalist terms. And the alternative "mentalistic" scheme is somehow to create a problem for the scientist.

Well, we cannot interpret the problem as involving a demand that the physicalist be able to give an explanation for "everything" for which an explanation may be desired. For we cannot seriously expect a scientific theory which explained truths of arithmetic. (And if we did the scientific theory would have to be incomplete anyway, since there are uncountably many arithmetical truths and only countably many possible explanations for them.)

Suppose the problem involved a demand for a separate explanation by the scientist to fit each possible description of the events to be explained. It might be supposed that this is as unreasonable a demand as one requiring a mathematician to prove arithmetical truths not only

in decimal notation, but also in binary notation, ternary notation, and so on. But in this mathematical case there is a strict and precise engagement between proofs in binary, ternary and so on: these could not be set up as rival proofs or explanations, since they are logically connected and complementary. Clearly, the present problem is focussed on the lack of engagement between the alternative "mentalistic" and physicalist schemes - a lack of engagement which may appear to set them up as rivals.

There is also a lack of engagement between a physicalist explanation for a flash of lightning and an explanation for one of Zeus's thunderbolts in terms of Zeus's whims. But in this case we can expect the scientific theory to discredit its rival and exhibit it as an inadequate or unreasonable account. But in the present case the "mentalistic" scheme is perfectly reasonable, and it would be highly undesirable for the physicalist account to discredit it even if it could do so, which is very unlikely. In other words, the physicalist scheme may appear threatened by an irreconcilable rivalry between it and the "mentalistic" scheme.

But the alternative "system" of mental discourse cannot in any general way be construed as a rival to the physicalist scheme. Nor can it be construed as a rival theory. Nor can the relationship between the two systems of explanation be regarded as scientifically mysterious, or mysterious at all.

Certainly, as I will mention, in particular cases there may be a rivalry or conflict between the two types of explanation. But then in particular cases it is also possible to engage the two types of explanation.

Consider the following example, not of human behaviour, but of plant "behaviour". Suppose someone asks 'Why is that plant growing towards the window?', and an explanation is offered: 'It's trying to

find more light'. An investigation of the detailed physical processes involved in the growth of the plant would yield in principle an explanation in terms of physical causal antecedents. And although the form of the latter explanation would be very different from that of the informal intentional explanation, it will clearly engage with it: the physicalist explanation is likely to confirm or refute the intentional explanation. But the engagement in this particular case may not involve any strict generalisable entailments between the two explanations. The intentional explanation would have been based on gross tendencies regarding plant growth, but may easily be reconciled with a detailed account in terms of causal antecedents, even though the intentional explanation is neutral as regards any prescription of matters of physical detail which figure in the physicalist explanation. At least intuitively we can see that it may not be difficult to compare the two explanations for compatibility.

Similarly, suppose someone is asked 'Why did A decide to attend that lecture?', and he replies: 'Because A was interested in what the lecturer had to say'. In principle a scientist may be able to pin-point as the principal cause of the chain of neural events which led to A's attendance, a neural memory trace of a conversation A had had with a girl; as a result of which the scientist may be able to refute the intentional explanation given, and suggest instead that A was interested in meeting the girl again at the lecture, and this was why he decided to attend it. Or, if the scientist confirms the explanation given, this need in no way trouble the scientist: for though the basis on which evidence for the intentional explanation was obtained is very different from the basis for the scientist's evidence, it is in general a sound basis - as good a basis as we can have if we cannot examine the causal antecedents involved - without being scientifically mysterious, and without <u>generally</u> rivalling the scientific basis. Though as we see there can be shown to

be possible conflicts in particular cases.

We should note also that the scientist would have available a very flexible hierarchy of concepts with which to specify neural processes: ranging from "pure" neural-descriptive concepts through "mixed" specifications like 'a neural memory trace of ...', to concepts in ordinary mental discourse.

The case here is assumed therefore to be rather like the case where a scientist can not only give a physicalist explanation of the events in a chess machine: he can also play chess, and is therefore able to move from one type of account to the other without difficulty. In other words, regarding explanations of human behaviour, we assume that the scientist can speak the natural language, and is acquainted with the concepts used in mental discourse to interpret human behaviour: he can, for example, not only describe the contortion of a human face in physiological terms - he is also familiar, having learnt the natural language, with the fact that such contortion is interpreted as "an expression of anger", for example. Clearly we have here a notion of "interpretation" which the pure physicalist account will try to avoid: just as a syntactical description of an output of the chess machine will avoid interpreting it as a move in chess.

This notion of interpretation of human behaviour is applicable in principle to the behaviour of any structure (as I noted in 5.22). But the more "complex" the behaviour, at least, the less "random" it is - the more it displays regular tendencies, structure and so on - the more susceptible it is to interpretation. Thus, if a certain machine's behaviour is such that it can be interpreted as answering arithmetical questions correctly, then we can describe and explain the machine's outputs according to this interpretation. For in interpreting the behaviour we implicitly assign and project a persisting structure to

the behaviour. We can then "forget about" the physical/syntactic "derivation", since we assume it fits the interpretation. But if the machine's outputs cannot be interpreted easily, then we are more likely to have to fall back on the physical/syntactical description and explanation: in this analogy (essentially Davidson's), a physicalist description and explanation. In the case of human behaviour, not only is it very susceptible to interpretation, but also the "syntactic derivations" are contingently hidden, so we only have available the interpretational account.

But as I noted in 5.22, we essentially cannot restrict the notion of conceptual dualism as it is manifested in possible descriptions of human behaviour, as just applicable to human behaviour. It is therefore not a special feature attached to our notion of human mentality. We must therefore expect the sort of conceptual dualism which may be supposed to be a crucial and problematic concomitant of human mentality, to be applicable in continuous degrees to structures of varying complexity and behaviour.

As I have tried to indicate, we may be able to render the conceptual dualism as harmless to scientific theory. But even if we can't, it cannot readily be construed as a special problem generated by our notion of the human mind.

So the possibility of ontological monism, and thus the possibility of a scientific explanation for the occurrence of all events, is the crucial possibility allowing as in principle possible a scientific accommodation of the human mind: the human mind does not then present a special logical obstruction to scientific treatment. So through the augmented Identity Theory we can endorse physicalism in its strong form.

5.3 Problems of Unity and Divisibility

Physical structures as we conceive them are physically divisible in an arbitrary manner. I will briefly discuss two related sorts of problem which this divisibility may be taken to pose for the augmented Identity Theory: essentially arising from notions of the "unity" or "simplicity" of mind.

5.31 A physical structure is in a sense only arbitrarily regarded as "one" structure. Certain events which could occur in a physical structure could not occur if the structure was physically divided; but many events occurring in the structure may still be able to occur even if the structure is physically divided. Our central nervous systems must be subject to this possibility.

Certainly, we should expect that some types of neural event could only occur in an intact central nervous system. But we cannot readily expect any other sense in which the nervous system could be regarded as "unified".

In contrast, there is an intuitive tendency, and a philosophical tendency, to regard "the mind" as indivisible, as not just arbitrarily "one". As Nagel says, we have 'the idea of a single person, a single subject of experience and action' (Nagel (2), p. 111).

There is a tendency to regard one's "stream of consciousness" as not just the contemporaneous and logically independent occurrence of distinct events. There seems perhaps a sense in which one's sensations, thoughts, and so on, are tied together by one subject of awareness. And although it would be claimed that one's stream of consciousness is in a sense divisible conceptually, in that the subject can introspectively recognise parts of the whole of which he is aware, it would be claimed that that is the only sense of division applicable to consciousness:

there could be no independent existence for thoughts, visual, aural sensations, and so on, in which they were not embraced by a common subject.

Even so, a person who claims this would have to accept some ontological independence: if one is blinded, one's visual imagings cease to occur, while "the rest carry on". But a person might suppose that there couldn't be a sense in which his visual imagings occurred and he didn't "know about them". Yet if visual imagings are neural events we must suppose that in principle such events may occur "in isolation" - accompanied by no other neural/mental events; or that visual imagings may occur but artificially isolated from the rest of the neural stream.

In Part 4 I endorsed a view in which these must be genuine possibilities, and in which the tendency to regard the neural/mental stream as peculiarly "unified" was opposed (cf. 4.42, (5) and 4.52). And on the augmented Identity Theory we should expect that if the brain was physically divided in a way which allowed the divided parts to sustain independent or partially independent functions, then independent neural/mental streams would result.

Nagel discusses cases of brain bisection in which the "higher" connections between the two cerebral hemispheres have been severed in men, monkeys and cats (Nagel (2)). And Nagel's discussion complements the present view, for he finds that there is great difficulty in persisting with the notion that the patient whose brain is bisected should be considered as possessing either one "individual mind" or any clear number of individual minds. Nagel concludes:

> If I am right, and there is no whole number of individual minds that these patients can be said to have, then the attribution of conscious, significant mental activity does not require the existence of a single mental subject. ((Nagel (2), p. 124.))

On the present view we must envisage not only this sort of malleability of "personhood" but also the malleability suggested by Parfit regarding the notion of a single person enduring through time. Mental/neural events just occur contemporaneously without being embraced by the awareness of a (ghostly) subject; and just occur successively without this succession being embraced by the awareness of a constant persisting subject.

This is brief. But at any rate we see the sort of line which the augmented Identity Theory would follow in order to parry objections based on the unity of consciousness.

5.32 The problem I will now describe involves the apparently finitary nature of mental events. It is not a problem I find easy to express; but if it is coherently expressible, it is perhaps the most difficult problem which the augmented Identity Theory has to overcome. (My presentation of this problem and its possible solution is rather compressed and difficult. But I hope it indicates at least an outline of some of the issues involved and of how the problem might be solved.)

Consider an abstract Turing machine. The logical states of such a machine are conceived as finite and discrete, in much the same way as formulae in finitary formal systems are conceived as finite and discrete - they involve a finite number of discrete "components". Moreover, the passage of an abstract Turing machine through its logical states is conceived as a passage through a finite number of discrete states.

When we consider a physically realised Turing machine, its finitary nature is reflected in this way: although the passage of the Turing machine through its states is a continuous physical process, involving (we assume) an infinitely divisible number of changes, there are only a finite number of "relevant" changes in the physically realised

machine. If we considered, for example, a machine constructed out of thermionic valves, there would be (we assume) a continuous and infinitely divisible physical change in an individual valve, and an infinite number of indiscrete states involved in that change; but the states which are relevant to the logical state of the machine are the valve's being "on" and the valve's being "off".

Moreover, there may be many other physical and therefore continuous changes in the physically realised machine which are irrelevant as far as the changes we're interested in are concerned: for instance, the machine may change in temperature slightly without affecting the relevant operations.

Similarly, a very slight change in temperature in a neural structure, or an event in an individual neurone too slight to cause a neural "ripple" to spread through the neural structure, would both be <u>neural events</u>. But <u>these</u> events we may not wish to regard as relevant to the functioning of the brain.

Thus we must suppose that there are an infinite number of continuous changes constituting neural events, even though only a finite number of those changes may be considered relevant. So, though we may in principle be able to specify the discrete states of a neural structure which we wish to regard as relevant - in much the same way we could specify the discrete states of a physically realised Turing machine - what <u>actually occurs</u> in the neural structure during its passage through these discrete and essentially abstract states is a continuous and (we assume) infinitely divisible physical process.

Now, let's take the case of visual imagings. It will be convenient for exhibiting the present problem to consider a "visual image" in a sense abstracted from 'visual imagings': a visual image in this sense we may roughly take as a visual imaging "frozen" at a particular point

in time. (In 4.41, (8) I identified a visual image in this sense with an open wff of the appropriate temporal J-language.)

Now, it may be argued that a visual image comprises only a finite number of "parts" - even though we are not introspectively aware of any "discontinuity". (It is essentially because of the "continuity" involved that I applied the notion of a graphic J-language to this case.) But despite this "continuity", we cannot readily suppose that a visual image is infinitely divisible. (This manifests itself, for example, in there being only a finite number of colour discriminations we can make, even though our colour spectrum does not appear to have any "gaps" in it.)

Thus we assume that a visual image, construed as an open wff of the appropriate J-language, maps only a finite number of features of the "frozen" event which it J-designates. In respect of its finitary nature, the visual image seems to correspond to what we would consider as the logical state of a physically realised Turing machine. (Indeed, I suggested in 4.32, (11c) that we may treat an open wff of a temporal J-language rather like a logical state of a Turing machine.[12])

It seems then that a sequence of visual imagings, construed as a sequence of visual images, is a finite sequence of finite entities. So it may seem, very oddly, that what we want to say on the augmented Identity Theory actually occur are very much like abstract entities. And it might be supposed that however odd this is, we cannot therefore say that mental events are neural events, since mental events seem to be finitary and neural events seem not to be. (And even if ultimately physics were to suggest that there are only a finite number of things or events, there would still be more changes going on in a neural event than in a mental event.)

One approach to this problem is essentially Leibnizian. This approach

is undesirably metaphysical, but I will mention it out of interest.

Leibniz's introduction of monads was motivated in part by a very similar problem connected with infinite divisibility. Suppose we take a certain sort of physical object as a real and ultimate constituent of the world. Well, we would in principle be able to subdivide such objects into components. And relative to these components, the objects we specified first were only arbitrarily regarded as individual objects. There isn't the first sort of object and its components arranged to constitute it. But we could subdivide any physical object - so for any sort of physical object we might take as the ultimate constituent of the world, there would be constituents of it which would render it not an ultimate constituent. So the ultimate constituents of the world could not be physical objects. So Leibniz introduced finite discrete entities - monads - as the real constituents of the world.[13] So the world would then contain an infinite number of finite monads.[14]

Now, the present problem is amenable to a similar "solution". Suppose we start from the assumption that a visual imaging J-designating an event E actually occurs (is given ontological status), but is finite. Then to get back to monism we must remove from the ontology infinitely divisible events. So we say that what suggests to us that E is infinitely divisible is merely that there are in principle an infinite number of possible finite J-designations of E - an infinite number of finite entities which "mirror" E. And then we remove E from the ontology by construing it as an abstraction,in a similar manner to the removal of physical objects from Leibniz's ontology. On such a view the world would comprise an infinite number of discrete changes.

A more satisfactory approach to the present problem is this. We simply deny that mental events are finitary.

We can allow, for example, that visual images, as I have construed

them, are composed of a finite number of "parts". But then visual images are abstractions. What actually occurs, the sequence of visual imagings, is infinitely divisible: it is infinitely divisible in "time". Of course we don't notice, in the sense of being able to count off, an infinite number of "temporal" positions over which a visual imaging extends - nevertheless, a visual imaging extends over an infinite number of "temporal" positions.[15]

So the objection that there are more "parts" to neural events than to mental events cannot be sustained. In the case of visual imagings, the finitary element involved adheres only to the open wffs of the appropriate J-language - the visual images. The closed wffs, of which we can still say there are a finite number in a sequence, have an infinitely divisible extension.

Clearly there are issues involved to which I cannot do justice here - but I hope it's possible to see how the mind-body problem involved here may not be as intractable as it may appear at first sight.

5.4 Conclusion

The Wittgensteinian line of argument which I drew in Part 2 suggests as incoherent a certain sort of mind-body problem involving ontological dualism sustained by a notion of privacy. The difficulty with the Wittgensteinian line is that it seems as though it can only dissolve mind-body problems if human beings are construed as "automata" whose mental life is just a "blank".

The Identity Theory, essentially conceived as a more palatable alternative to "blank automatism", seemed either to refloat the human mind as an entity exposed by Wittgenstein as unstable; or else do little to alleviate the apparent consequence of the Wittgensteinian line.

But through the augmented Identity Theory we avoid a view of "blank automatism" while acceding to the restrictions imposed by the

Wittgensteinian line - essentially by drawing the realist distinction in a way which fits the Wittgensteinian picture.[16]

I have suggested that once ontological monism is seen as unobjectionable, the difference between ordinary mental discourse and the terminology envisaged in a physicalist theory presents no serious logical difficulty for the scientist's accommodation of the human mind.

But one concept attached to mental discourse which may appear especially to strain this suggestion is the concept of freedom of action. In at least the following sense the idea that we act freely can be taken as incompatible with physicalism: the degree and nature of any indeterminism allowed in a physicalist view cannot accommodate the intuitive view that we can in a deep sense help what we do, make voluntary decisions, and so on. For the notion of indeterminism which may be allowed within physicalism seems applicable only to the micro-world in a way which cannot account for the indeterminism which is at least apparently required for the intuitive view of freedom. And in any case, indeterminism in physics - at least at present - is associated with randomness; and this is hardly the sort of notion which could accommodate the intuitive notion of voluntary action.

Clearly it is of the utmost importance for a physicalist view to exhibit how the concept of freedom works in a manner compatible with there being constraining causal antecedents for our behaviour; and similarly to exhibit the workings of explanations for actions in mental discourse, which may seem at least superficially to be at odds with the view that actions could be explained in terms of causal antecedents.

But we can see at least that the fact that we <u>do</u> explain our behaviour in the way we do is quite compatible with our being purely physical and fully constrained causally. For purely physical structures

which were fully constrained causally could behave just as we do - in particular, if they had only a minimal knowledge of their causal constraints, they could and indeed would have to explain their behaviour in the way we do ours.

If physicalism has to curtail the intuitive notion of freedom of action - and I think it does have to - then, ironically perhaps, the resulting view of human beings allows a rather more compassionate and humane view of individuals than we appear to have a present.

NOTES TO PART 5

[1] 'Some writers ... have used the word matter to mean the phenomenon plus the reality represented ... But for the purposes of the present discussion I have thought it clearer to use the word for the phenomenon as distinguished from the thing-in-itself.' (Ibid., p. 73).

[2] Clifford was by no means entirely naive about the problems of meaning involved: he mentions the notion of a "social object", in the case of the table: '"this table as an object in the minds of men"' (Ibid., p. 56), in contrast to the "individual object" in one person's consciousness; and he says interestingly:

> ... any sound which, becoming a sign to my neighbour, becomes thereby a mark to myself, must by the nature of the case be a mark of the social object, and not of the individual object. ((Ibid., p. 57.))

[3] Clifford's earlier view (cf. Clifford (1)) was a sort of parallelist panpsychism; and in his later paper he inconsistently lapses into his earlier view when he allows a sense in which there can be said to be a "mental component" of matter:

> A moving molecule of inorganic matter does not possess mind or consciousness; but it possesses a small piece of mind-stuff. ((Clifford (2), pp. 69 - 70.))

[4] This is not quite the illegitimacy mentioned in the citation: the illegitimacy of supposing (roughly) "mind" to be of the same kind as "matter". The other illegitimacy or incoherence involves the supposition that (roughly) "mind" is of the same kind as things-in-themselves.

[5] But it should be clear that on the present view we should avoid saying with Schlick: 'The perceptions themselves are reality experienced'. (My underlining.)

[6] Thus different observers might disagree about the distance between two events, or the time gap between two events - but all observers would

agree about the "interval" between two events, which is given by a deducible mathematical function of both spatial and temporal separations. An analogy for the relation between the relativistic interval and spatial and temporal separations can easily be given. Suppose we have a straight line. We can regard the distance between any two points on the line as analogous to the relativistic interval. Now, if we frame the line in a particular Cartesian co-ordinate system, the distance or interval between two points on the straight line reflects itself in the projections (of the segment of the line between the two points) on the X and Y axes: the projections being analogous to spatial and temporal separations. The projections would depend on which Cartesian frame is chosen. But there is a mathematical function which yields the "objective" interval if we feed in the magnitude of the projections, whatever the Cartesian frame: $\sqrt{x^2 + y^2}$, where x and y are the magnitudes of the projections on the two axes.

[7] On the assumption that mental events are neural events, if my skull merely contained a device which transmitted my sensory nerve impulses to Sirius, where my brain was housed, equipped with a receiver and transmitter, I would in fact notice that my bodily reactions were somewhat "delayed". If my brain was similarly housed somewhere on Earth, I might not notice the slight "delay".

[8] Eddington speculatively mentions the issue of Relativity and the brain in Eddington, pp. 180 - 201. He also endorses Clifford's view.

[9] The neatest case for this sort of comparison is where we have two different physical realisations of one abstract Turing machine. Though the detailed physical mechanisms involved may be intrinsically quite different, there is a clear sense in which the gross behaviour of the two machines - in this case their outputs - could be used to compare extrinsic features of the detailed mechanisms involved.

[10] It may have been noticed that now that we cannot specify mental events as being in principle private, it is not quite as easy to delineate exactly what specifications couched in "mental discourse" specify the events which on the augmented Identity Theory are neural events. The best I can do, and indeed given the nature of the present strategy, which is intended to leave an ontological monism of physical events of varying degrees of complexity with no intrinsic physical property picking any out as "mental", perhaps the best we can expect, is this:- we recognise the class of events which may be *prima facie* problematical to any rich enough physical structure: those events for whose occurrence the structure has evidence, *even though* they may be contingently imperceptible both directly and indirectly to the structure. In other words, mental events in a human being are those events for whose occurrence he can be said to have (introspective) evidence, and which can be reported by him as occurring *even though* they are *contingently imperceptible* to him both directly and indirectly: roughly, events which can be said to be part of his "stream of consciousness". Clearly it would go against the essential character of the augmented Identity Theory to include indiscriminately as mental events figuring in the Identity Hypothesis just *any* events whose specification is "couched in mental discourse". Since such events may not be of the sort we can be introspectively aware of, and for which we have not (at present) a physical description:- we could assert (in "mental discourse") the occurrence of events which would be construed as "segments of behaviour", for which we could have a neutral non-interpretational physical description without needing to describe *neural* events - for example 'I accepted his challenge on the spot'.

[11] This is of course present in Davidson's account.

[12] We should note that a combination of discreteness and "continuity" is also retained in the notion of logical states of a Turing machine. For though of course there is a "discontinuity" in the symbolic representations of logical states, the discreteness of the structure of a logical state cannot be regarded as discontinuity in the sense of there being a "gap" between the elements 0 and 1 determining the structure of the logical state. There is of course a discontinuity between '0' and '1'; but not between the logical states: for the notion of a discontinuity requires a medium in which we can pick out a separation.

[13] Though Leibniz was not just motivated by the considerations rendered here.

[14] We should note that it was not the notion of there being an infinite number of things which bothered Leibniz, but that the things should be genuine things, which for him meant that they should have definite "edges". In analogy note the slight discomfort which may be experienced in connection with the notion of a real number (as opposed to a natural number): we may only be able to specify a real number by an endless row of decimals, in which case we could not finitely specify it.

[15] I use '"time"' and '"temporal"' rather than 'time' and 'temporal' to alert ourselves to the possibility, voiced in 5.13, that we should not strictly construe the ordering relation on visual imagings as temporal if we are to accede to relativistic considerations. We can use relativistic considerations to parry the objection that neural events are infinitely divisible in time and space. (If we recall an analogy given earlier: an infinite number of points on a line framed in a Cartesian co-ordinate system project themselves as an infinite number of points on the two axes. But the line itself comprises just one set of points.) All we need is for a sequence of visual imagings to be infinitely divisible: which it is.

[16] The realist distinction is given by Kant in a very complex and rather strained form. It may be possible to construe part of my line of argument as an attempt to remodel certain elements of the Kantian view along lines which avoid the confusions exposed by Wittgenstein. It would I think be useful and feasible to attempt this in a more explicit manner.

REFERENCES

Abbreviations for names of journals:

A	Analysis
AJP	Australasian Journal of Philosophy
APQ	American Philosophical Quarterly
D	Dialogue
JP	Journal of Philosophy
M	Mind
P	Philosophy
PAS	Proceedings of the Aristotelian Society
PASS	Proceedings of the Aristotelian Society, Supplementary volume
PQ	Philosophical Quarterly
PR	Philosophical Review
RM	Review of Metaphysics

Abbreviations for collections of essays:

Anderson	A.R.Anderson (ed.) Minds and Machines, Prentice-Hall, Englewood Cliffs, 1964.
Borst	C.V.Borst (ed.) The Mind-Brain Identity Theory, Macmillan, London, 1970.
Hampshire	S.Hampshire (ed.) Philosophy of Mind, Harper and Row, New York, 1966.
Hintikka	J.Hintikka (ed.) The Philosophy of Mathematics, Oxford University Press, 1969.
Jones	O.R.Jones (ed.) The Private Language Argument, Macmillan, London, 1971.
Pitcher	G.Pitcher (ed.) Wittgenstein: The 'Philosophical Investigations', Macmillan, London, 1968.

Vesey — G.N.A.Vesey (ed.) Body and Mind, Allen and Unwin, London, 1964.

A date in brackets after the title of a work is the date of its first publication; and if the name of a journal also occurs within those brackets, it is the journal in which the work was first published.

Austin, J.L. — Sense and Sensibilia, Oxford University Press, 1962.

Ayer, A.J. — 'Can There Be a Private Language?' (PASS, 1954) in Jones, pp. 50-61.

Armstrong, D.M. — 'The Nature of Mind' (1965) in Borst, pp. 67-79.

Baier, K. — 'Smart on Sensations' (AJP, 1962) in Borst, pp. 95-106.

Boring, E.G. — 'Mind and Mechanism', American Journal of Psychology, vol. 59, 1946, pp. 173-192.

Bradley, M.C. — 'Two Arguments against the Identity Thesis', in Contemporary Philosophy in Australia, (ed.) R.Brown and C.D.Collins, Allen and Unwin, London, 1969, pp. 173-189.

Brentano, F. — Psychology from an Empirical Standpoint (1874), (ed.) L.L. McAlister, (trans.) A.C. Rancurello et al., Routledge and Kegan Paul, London, 1973.

Castaneda, H-N. — 'The Private Language Argument' (1962) in Jones, pp. 132-154.

Clifford, W.K. (1) — 'Body and Mind' (Fortnightly Review, 1874) in Lectures and Essays, W.K.Clifford, Macmillan, London, 1901, vol.2, pp.1-51.

(2) — 'On the Nature of Things-in-themselves' (M, 1878) in Lectures and Essays (see Clifford (1)), vol.2, pp. 52-73.

Cornman, J.W. (1) — 'The Identity of Mind and Body' (JP, 1962) in Borst, pp.123-129.

(2) — 'Private Languages and Private Entities', AJP, vol.56, 1968.

Davidson, D. (1) — 'Mental Events', in Experience and Theory (ed.) L.Forster and J.W.Swanson, Duckworth, London, 1970, pp. 79-101.

(2) 'The Material Mind', in Logic, Methodology and Philosophy of Science IV (ed.) P.Suppes et al., North Holland, Amsterdam, 1973, pp. 719-722.

Dennett, D.C. — Content and Consciousness, Routledge and Kegan Paul, London, 1969.

Descartes, R. — 'Meditations on First Philosophy' (1641) in Discourse on Method and Other Writings (trans.) F.E.Sutcliffe, Penguin, Harmondsworth, 1968.

Donnellan, K.S. — 'Reference and Definite Descriptions', PR, vol.75, 1966, pp. 281-304.

Dummett, M. (1) 'Wittgenstein's Philosophy of Mathematics' (PR, 1959) in Pitcher, pp.420-477.

(2) Frege: Philosophy of Language, Duckworth, London, 1973.

Eddington, A.S. — Space, Time and Gravitation (1920) Harper and Bros., New York, 1959.

Feigl, H. (1) The "Mental" and the "Physical": The Essay and a Postscript, University of Minnesota Press, Minneapolis, 1967.

(2) 'Mind-Body, not a Pseudo-Problem' (1960) in Borst, pp.33-41.

Frege, G. (1) The Foundations of Arithmetic (1884), (trans.) J.L.Austin, Blackwell, Oxford, 1953.

(2) 'On Sense and Reference' (1892) in Translations from the Philosophical Writings of Gottlob Frege (ed. and trans.) P.Geach and M.Black, Blackwell, Oxford, 1960, pp. 56-78.

(3) 'The Thought: A Logical Enquiry' (1918), (trans.) A. and M.Quinton, in Philosophical Logic (ed.) P.F.Strawson, Oxford University Press, 1967, pp. 17-38.

Gardner, M. — The Ambidextrous Universe (1964) Penguin, Harmondsworth, 1970.

Geach,P.T. 'Some Problems about Time' (1965) in Logic Matters, P.T.Geach,Blackwell,Oxford,1972,pp. 302-318.

Glover,J. (ed.) The Philosophy of Mind,Oxford University Press,1976.

Hervey,H. 'The Private Language Problem' (PQ,1957) in Jones,pp.76-95.

Hinton,J.M. 'Illusions and Identity' (A,1967) in Borst,pp.242-257.

Hopkins,J. 'Wittgenstein and Physicalism',PAS,vol.75,1975,pp.121-146.

James,W. The Principles of Psychology (1890) Macmillan,London, 1901,vol.2.

Kant,I. Immanuel Kant's Critique of Pure Reason,(trans.) N.K.Smith, Macmillan,London,1929.

Kenny,A. 'The Verification Principle and the Private Language Argument',in Jones,pp. 204-228.

Kneale,W.C. & M. The Development of Logic,Clarendon Press,Oxford,1962.

Kripke,S.A. 'Naming and Necessity' in The Semantics of Natural Language, (ed.) G.Harman and D.Davidson,Dordrecht,1972,pp.253-355.

Lucas,J.R. 'Minds,Machines and Godel' (P.1961) in Anderson,pp.43-59.

Malcolm,N. 'Scientific Materialism and the Identity Theory' (D,1965) in Borst,pp. 171-180.

Mundle,C.W.K. '"Private Language" and Wittgenstein's Kind of Behaviourism' (PQ,1966) in Jones,pp. 103-117.

Nagel,T. (1) 'Physicalism' (PR,1965) in Borst,pp. 214-230.

(2) 'Brain Bisection and the Unity of Consciousness' (Synthese, 1971) in Glover,pp. 111-125.

(3) 'What Is It Like to Be a Bat?',PR,vol.83,1974.

Parfit,D. 'Personal Identity' (PR,1971) in Glover,pp.142-162.

Pears,D.F. Wittgenstein,Fontana,London,1971.

Place,U.T. 'Is Consciousness a Brain Process?' (British Journal of Psychology,1956) in Borst,pp. 42-51.

Prince,M. The Nature of Mind and Human Automatism (1885) extracts from, in Vesey,pp. 187-196.

Putnam,H. (1) 'Minds and Machines' (1960) in Anderson,pp. 72-97.

(2) 'Robots: Machines or Artificially Created Life?' (JP,1964) in Hampshire,pp. 63-91.

(3) 'The Mental Life of Some Machines' (1967) in Glover, pp. 84-100.

Quine,W.V.O. (1) Mathematical Logic (1940) Harper and Row,New York,1962.

(2) Word and Object (1960) Massachusetts Institute of Technology Press,1964.

Quinton,A.M. The Nature of Things,Routledge and Kegan Paul,London,1973.

Rhees,R. 'Can There Be a Private Language?' (PASS,1954) in Jones, pp. 61-75.

Rogers,H. 'The Present Theory of Turing Machine Computability' (1959) in Hintikka,pp.130-146.

Rorty,R. 'Mind-Body Identity,Privacy and Categories' (RM,1965) in Borst,pp. 187-213.

Rotman,B. and Kneebone,G.T. The Theory of Sets and Transfinite Numbers,Oldbourne, London,1966.

Russell,B. Human Knowledge: Its Scope and Limits,Allen and Unwin, London,1948.

Ryle,G. The Concept of Mind (1949) Penguin,Harmondsworth,1963.

Schlick,M. Allgemeine Erkenntnislehre (1925) extracts from (trans.) G.Brown,in Vesey,pp. 297-307.

Sellars,W. 'The Identity Approach to the Mind-Body Problem' (RM,1965) in Hampshire,pp. 7-30.

Shaffer,J. 'Could Mental States Be Brain Processes?' (JP,1961) in Borst,pp. 113-122.

Smart,J.J.C. (1) 'Sensations and Brain Processes' (PR,1959) in Borst,pp.52-56.

(2) 'Further Remarks on Sensations and Brain Processes' (PR,1961) in Borst, pp.93-94.

(3) 'Brain Processes and Incorrigibility' (AJP,1962) in Borst, pp. 107-109.

(4) 'Materialism' (JP,1963) in Borst,pp. 159-170.

Smullyan,R.M. 'Languages in which Self Reference is Possible' (Journal of Symbolic Logic,1957) in Hintikka,pp. 64-77.

Spinoza,B. The Ethics (1677) extracts from in Vesey,pp. 54-61.

Stevenson,J.T. '"Sensations and Brain Processes": A Reply to J.J.C.Smart' (PR,1960) in Borst,pp. 87-92.

Strawson,P.F. (1) 'Review of Wittgenstein's "Philosophical Investigations"' (M,1954) in Pitcher,pp. 22-64.

(2) Individuals,Methuen,London,1959.

Thomson,J.J. 'Private Languages' (APQ,1964) in Jones,pp. 183-204.

Turing,A.M. 'Computing Machinery and Intelligence' (M,1950) in Anderson,pp. 4-30.

Wittgenstein,L. (1) The Blue and Brown Books (1958) Blackwell,Oxford,1969.

I Philosophical Investigations (1953) Blackwell,Oxford,1968.

(3) '"Private Experience" and "Sense Data"' (PR,1968) (ed.) R.Rhees,in Jones,pp. 229-275.

Wundt,W.M. Lectures on Human and Animal Psychology,(trans.) J.E. Creighton and E.B.Titchener,Macmillan,New York,1901.