

EXPERIENCE, AGENCY AND THE SELF

Richard M Gaskin (St. Edmund Hall)



Thesis submitted in accordance with the regulations for the D.Phil.
examination of Oxford University. Trinity Term, 1988.

ACKNOWLEDGEMENTS

I should like to thank my supervisor, Professor Michael Dummett, for the generous assistance he has given me during the preparation of this thesis. I should also like to thank Professor Barry Stroud and Mr. David Wiggins for their helpful comments on earlier versions of several chapters. Lastly, I feel that an older debt of gratitude to Mr. John McDowell, who supervised the B.Phil Dissertation out of which the present thesis was developed, should not go unrecorded.

Richard M Gaskin, St. Edmund Hall, Oxford.

Title of Thesis: Experience, Agency and the Self.

Degree: D.Phil. (Oxford University)

Term of Submission: Trinity, 1988.

ABSTRACT

Wilfrid Sellars has made familiar a distinction between manifest and scientific images of man-in-the-world. The manifest image is 'a sophistication and refinement of the image in terms of which man first came to be aware of himself as man-in-the-world' ([2], p.18), and in its methodology 'limits itself to what correlational techniques can tell us about perceptible and introspectible events' (p.19). The scientific image, on the other hand, 'postulates imperceptible objects and events for the purpose of explaining correlations among perceptibles.' (ib.) This thesis is centred on a consideration of two difficulties facing anyone who takes the manifest image seriously as an autonomous image of man. In chapter 1 I consider the connection between perception and its objects, and argue that there is a disharmony between the manifest and scientific accounts of this connection. But I also suggest that the manifest image, which incorporates a certain Cartesianism or internalism, cannot lightly be dispensed with in our understanding of the nature of experience. Chapter 2 is a companion piece to chapter 1: in it I argue that the manifest view of experience accords a certain metaphysical priority to secondary over primary qualities in the constitution of any world capable of being experienced; I also

suggest that the scientific image is dependent on the manifest image, and so cannot subvert it. In chapter 3 I turn to the other main area of difficulty: freedom. I argue that free will as the incompatibilist construes it is constitutive of the time-order; but that it carries with it implicit internal contradictions. The conflict here lies within the manifest image; the scientific image discerns no such freedom, and so incurs no such problems. But if I am right that freedom constitutes time, it will not be an option for us to disembarrass ourselves of the contradictions. I also argue that there is a relation of mutual dependence between freedom, incompatibilistically construed, and internalism. The manifest image as a whole - deeply problematic as it is - is therefore grounded in and entailed by something quite ineluctable, namely the reality of the time-series. This is the principal conclusion of the thesis. If I succeed here, I provide support for the claim that our difficulties with the manifest image cannot be solved by abandoning it: the manifest image, problems and all, must just be lived with. The remainder of the thesis explores topics related to this main thrust. Chapter 4 is really an appendix to chapter 3; it shows how no parallel difficulties attend the constitution of experiential space, because space is (unlike time) not transcendental. In chapter 5 I examine the commitments of the notion of the transcendental self, whose existence was deduced in chapter 3 as a condition of freedom. In particular, I aim to show how that self inherits some of the difficulties of its parent concept of freedom; but also how a distinction between transcendental and empirical components in the self can help us with the problem of privacy.

CONTENTS

<u>1</u>	<u>A Theory of Perceptual Knowledge</u>	
1.1	Intentionality: a non-natural phenomenon	1
1.2	Error in Perception	18
1.3	The Internalist View	29
1.4	Appearances	39
1.5	Externalism and the Manifest World	49
<u>2</u>	<u>Objectivity and Secondary Qualities</u>	
2.1	The Causal Ground Argument	62
2.2	The Marks of Objectivity	75
2.3	The Status of the Primary	96
<u>3</u>	<u>Agency and Time</u>	
3.1	The Ineliminability of A-time	107
3.2	The Constitution of A-time: the Failure of Causality	120
3.3	The Constitution of A-time: the Need for Free Agency	133
3.4	Freedom and Causality: the Transcendental Status of the A-series	155
3.5	Freedom and the Transcendental Self	163
3.6	The Phenomenology of Freedom	178
3.7	Freedom and Internalism	187

<u>4</u>	<u>Experience and Space</u>	
4.1	Particularity and its Concomitants	196
4.2	The Constitution of Particularity	206
<u>5</u>	<u>The Transcendental Self</u>	
5.1	The Phenomenology of the Self	223
5.2	The Transcendental and Empirical Selves	233
5.3	The Identity of the Transcendental Self	240
5.4	The Question of Privacy	255
5.5	Vere's Choice	267
	FOOTNOTES	
	Chapter 1	278
	Chapter 2	289
	Chapter 3	293
	Chapter 4	306
	Chapter 5	308
	REFERENCES	318

Chapter 1: A theory of perceptual knowledge.

1.1 Intentionality: a non-natural phenomenon.

Much of our experience represents the world as being a certain way. What is the relation between the experience which represents and the world which is represented? It is common to suppose that causality must feature somewhere in the connection between the world and a representation of it. Another view is

that the correctness of the description of a perceptual experience as the perception of a certain physical thing logically requires the existence of that thing; and the logical is thought to exclude the causal connection, since only logically distinct existences can be causally related. (Strawson [4], pp.51-2)

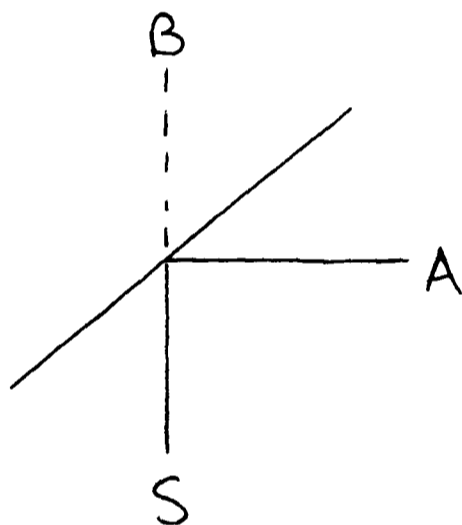
My starting point is my belief that the view alluded to in this passage is correct. What is meant by calling the connection logical is this: if I see a cup in front of me, the existence of the cup is analytically guaranteed by the perceptual status of my mental state: if the cup weren't there I wouldn't be either seeing it or seeing it. But causation is a contingent relation: if we cannot specify or give the description of an 'effect' without implying the existence of the 'cause', then it cannot be causation that is in question. Cause and effect are, as Hume said, logically distinct existences: what is meant by this slightly solecistic claim (since logical relations are normally thought to connect propositions, not things) is that a proposition stating the existence of a cause

neither implies nor is implied by a proposition stating the existence of an effect.¹ Upon this truth depends the a posteriori status of natural science, since the relations it examines are causal. Of course we can define words as we like, but if we want a uniform notion of causality which is fitted to service the project of constructing a natural science, it looks as if we should not try to locate it in the relation between perception and its objects.² We might say, echoing Brentano (2.1.5), that intentionality in perception is a non-natural phenomenon.

Echoing Brentano, but not exactly recapitulating him. Brentano was concerned with the world-directed content of thought in general, the fact not only that 'in presentation something is presented' but also that 'in judgment something is affirmed or denied, in love loved, in hate hated, in desire desired and so on' (ibid). What is common to all these modes of thought is their possession of content, and that was what Brentano meant by their intentionality. What I shall be concerned with - and what I shall mean by the word 'intentionality' - is the contentfulness specifically of perception-based thought. This species of thought has the peculiarity - this will be crucial in what follows - that the subject is always able to entertain de re or genuine singular thoughts about the objects presented to him in perception. In contradistinction, judges, desirers, haters and lovers do not universally enjoy such an intimate relation to the objects of their respective mental attitudes: in many cases the contents of these attitudes are purely general. My claim will be that the content of a perception is never purely general, and that the involvement of the

object in the perception rules out a causal relation between them.

In a well-known article³ Grice defends the location of causality between the world and experience in the following way (I alter some of the details): suppose a subject S sees an object A reflected in a mirror set at 45 degrees to the direction of his vision. Suppose further that he is not aware of the presence of the mirror, and that there is an object B, qualitatively indistinguishable from A, located at the position behind the mirror where S takes A to be located. The situation looks like this:



Grice's claim is that we need to bring in causal considerations to explain why S sees A and not B. Ex hypothesi there is no qualitative way to distinguish the experiences; the reason why S enjoys a perception of A and not B is that he is causally linked to A and not B.

Suppose we accept the Gricean point. That is, suppose we accept that the reason why S sees A and not B is that he is causally linked to A and not B. (Actually I shall qualify this acceptance in 1.2.) Does this mean that causality enters into the relation between

perception and its objects? I think not. For we are not here trying to explain why A rather than B is the object of S's perception of A: that needs no explanation, since B is not a candidate to be related to S's perception of A. We are trying to explain why S's mental state is a perception of A in the first place rather than a perception of B, and that is where it can be claimed we need causality. But once we have got as far as a description of S's mental state (a perception of A or an apparent perception of A) we have come too far for causality to have any role to play. There is no relevant causal connection between the manifest object (the object of S's experience), A, and S's mind; so the relation must obtain, when it does, between the object discerned by physical science (which I shall henceforth call the 'scientific object') with which A is correlated (and whose properties A's manifest properties supervene on) and S's brain.⁴ If we try to locate the effect in S's mind, we run up against the fact that A is an intentional object of S's mental state, and so cannot be contingently connected with it. This connection between A and S's mental state obtains even if that state is characterized as a state of seeming to see A rather than a state of seeing A. That is because there are not two states in question; and there are not two objects - a merely intentional one (the one S seems to see), and a real one (the one S sees). If S perceives A, then his state of seeming to see A is a state of seeing A; and the intentional object of his apparent perception is not merely an intentional object: it coincides with the real object, A, which is the object of his perception. It might be claimed that it is in virtue of a causal relation between A and S's mind that the

intentional object and the real object do coincide. But the difficulty remains: what is the effect in this causal transaction? Specify it. You cannot do so without implying the existence of the 'cause'. The mental state you refer to - however you make the reference - essentially involves the 'cause', the object A.

The view I am advancing obviously depends on a certain essentialism: the description of S's mental state as a perception of A is privileged in that it specifies the essence of that mental state. That mental state could not have failed to be a perception of A. It follows that there is an essential connection between A and that mental state, however these items are referred to. This view contrasts with one according to which a mental state is a perception of A 'only under description'; in that case contingency or necessity attaches only to the proposition stating the relation between A and S's mental state according as they are picked out under one description or another.⁵ Contingent and necessary propositions would be artefacts of ways of describing things, and would not identify features of things essentially or accidentally inhering in the things themselves. But this view imports an intolerable scheme/given dualism into our metaphysic.

Of course, no one would deny that, for example,

(A) The cause of the First World War caused the First
World War

is, at least on one construal, necessarily true, whereas

(B) The assassination of Archduke Ferdinand caused the First
World War

is only contingently true. But it does not follow from this that

there was one event which, characterized in one way, necessarily caused the First World War and which, characterized in another way, only contingently caused it. It does not follow that from talk of the necessity/contingency of the propositions (A) and (B) nothing can be recovered concerning the essential and accidental properties attaching to things. For the two characterizations in play - 'the cause of the First World War' and 'the assassination of Archduke Ferdinand' - are not on a logical par. The former is a definite description, and is contingently instantiated by the assassination; the latter is a proper name, and so necessarily attaches to its referent. That is: there is no necessary connection between the sense of a definite description and the object (if any) which instantiates that description; but the sense of a proper name necessarily presents the object which is the referent of that name (and there necessarily is such an object if the name has sense). One way of conveying the point is this: there is a way of construing (A) under which it comes out only contingently true. This involves reading the phrase 'the cause of the First World War' as a descriptive name, i.e. we hear (A) as saying

(A') Whatever in fact caused the First World War caused the
First World War

the contingency of which is indicated by the falsity of

(A'') Whatever in fact caused the First World War
necessarily caused the First World War

But there is no such variety of readings of sentences containing the phrase 'the assassination of Archduke Ferdinand'. Although it has the surface form of a description, it is in logical

grammar a proper name; the same goes for 'S's perception of A'. With respect to modal contexts, the distinctive feature of proper names is that they lack scope. S's perception of A could not have failed to be S's perception of A; whereas what caused the First World War could have failed to cause it. There is then a necessary connection between A and S's perception of A: the latter cannot exist unless the former exists. This is a necessity at the level of objects, not merely at the level of language. We can start off at the level of language if we like, and say that the proposition that S perceives A entails the proposition that A exists. But we cannot stop there: given that 'S perceives A' is equivalent to 'S's perception of A exists', and given that 'S's perception of A' is a proper name, it follows that S's perception of A essentially involves the existence of A. There is a necessary connection at the level of objects; so that the purported 'cause' and 'effect' are not logically distinct existences.

Formally, we should represent the necessity in question using the de re modal predicate 'NEC', not the sentential modal operator '□'.⁶ This is because the relation in question (that between the mental state and its intentional object) is asymmetric. If A is the intentional object of mental state M, then M is essentially such as to be of A, but A is not essentially such as to figure in M; M could have had no other intentional object than A, but A might not have figured in M (or in any mental state). If we symbolize the relation 'x has y as intentional object' by 'Of(x,y)', then the claim is that we do not do best to express the situation as:

$$(x)(y)(Of(x,y) \longrightarrow \square Of(x,y))$$

which says that if any mental state has an intentional object then necessarily: that mental state has that intentional object. This de dicto formulation is unsatisfactory because it is ambiguous between on the one hand the false

$$(x)(y)(\text{Of}(x,y) \longrightarrow [\text{NEC} [\lambda z(\text{Of}(x,z))]]], \langle y \rangle)$$

which says that if any mental state has an intentional object then that object has the property of necessarily figuring in that mental state; and on the other hand the true

$$(x)(y)(\text{Of}(x,y) \longrightarrow [\text{NEC} [\lambda z(\text{Of}(z,y))]]], \langle x \rangle)$$

which says that if any mental state has an intentional object, then that mental state has the property of necessarily containing that object.

This latter formulation is what we want to express the peculiar nature of intentionality, and it has no parallel in regular Humean causation. For the parallel to hold, we should need the truth of:

$$\exists x \exists y (\text{Causes}(x,y) \ \& \ [\text{NEC} [\lambda z(\text{Causes}(x,y))]]], \langle y \rangle)$$

which says that there is at least one thing such that it is essentially the effect of what it is an effect of. This is one of the possibilities which regular Humean causation excludes. In the case of intentionality, the asymmetry is precisely parallel to that involved in set-membership or the part-whole relationship: if, for example, $x \in y$, x does not have the property of necessary membership of y , but y does have the property of necessary possession of x as a member. (No one would suggest that a set was caused by its membership, or a whole by its parts.)

Consequently, to find a causal transaction between an object of perception and the perceptual mental state would break the Humean

requirement in one direction: it would make the cause logically non-distinct from the effect (although not the effect logically non-distinct from the cause). But one direction is enough. Ordinary causation is fully Humean in that relations of essential dependency obtain in neither direction.

It is true that cause and effect do, contra Hume, enjoy general a priori or necessary relations. If an object is solid it will in general exclude other objects; if it is brittle it will in general break when dropped, and so on. The discerning of such causal dispositions is fundamental to our ability to find material objects in the world. Of course science may come to dispense with any particular set of dispositions (solubility and fragility may be replaced by explanatorily more illuminating characterizations); but it cannot in principle dispense entirely with dispositions - at least not as long as it works with objects which may have spatial, but not temporal, parts.⁷ For such objects, it will always be the case that what it is to be an object of a certain sort involves the possession of dispositions, operative over time, to behave in certain ways in certain circumstances. But these a priori truths connecting objects with causal dispositions do not and cannot suffice for necessary connections between particular causes and particular effects. It would be fallacious to conclude from 'X is brittle at time t' and 'X is dropped at time t' to 'X breaks at time t'; or from 'Y is soluble at time t' and 'Y is deposited in water at time t' to 'Y has dissolved by time t+n'.⁸

It is obvious that such phenomena as sun-burn or sea-sickness are not genuine counterexamples to the claim about Humean causation:

the effects here can be perfectly well specified without bringing in the causes. We only use these designations of the effects because it usually is the sea which causes sea-sickness or the sun sun-burn. Were other causes to become more frequent (as is quite conceivable), we might see fit to change our names of the effects. The phenomenon of intentionality has no parallel in the natural world; the essential relation between perception and its objects, though asymmetric, is so striking that it warrants refusing to apply the terminology of causation to intentionality, at least if we want that terminology for regular Humean causation. It is simply the best way of registering the difference between the cases - the fact that intentionality is non-natural - to say that perceptions are not caused by their objects.

The alternative to the view I advocate is a metaphysic according to which there is a level at which S's mental state can be a determinate particular, but at which it is not yet a perception of A (or even an apparent perception of A). But the notion of such a bare particular is incoherent: S's mental state is a perception of A all the way down, and not 'only under description'. If not, then the bare particular which is in fact S's perception of A could have been a hat, say; which is absurd.⁹ More to the point, it could have been a perception of B, which is equally absurd; but in this case we need to say more precisely why.

What I claim is that there is no specification of S's mental state when he perceives A which ascribes to him anything less than a de re contact with A. De re thoughts differ from descriptive thoughts in the following way: they typically employ demonstrative

elements, or proper names, they enjoy a certain directness in the way they relate to their objects, and it is a defining feature of them that they are subject to the constraint that where there is no object of the appropriate sort, there is no thought of the essayed sort (see further 1.2 and 4.1). Now a causal relation between A and S's mental state cannot get going, I claim, because once we start attempting to characterize S's mental state, there is no stopping point on the road to ascription of a fully de re such state. There is a longstanding tradition in philosophy which supposes, on the contrary, that there is available an adequate specification of S's mental state which does not commit one to relating it intentionally to an external particular. If this tradition were right, there could indeed be a non-essential connection between the objects of perception and whatever mental state were offered as a surrogate of the full-blooded intentional state which relates to A.

We have already disposed of one candidate for surrogacy: S's apparent perception of A, or his perception of the merely intentional object, A. This surrogate failed to provide an adequate characterization of S's mental state, because the intentional object is identical with the real object: the object A plays both of these roles. Two other candidates have been commonly put forward to assume the role of the surrogate: unconceptualized sense-data, and perceptions 'as of'. The first of these has fallen out of favour in recent years, and for a very good reason. Michael Williams gives an excellent account of it: he objects that if the mind were presented, in the first instance, with a raw stretch of unconceptualized 'given', there would be no principled way for it to make the leap of

conceptualization. A subject would have to convert something with no propositional content (sheer non-conceptual phenomenology) into something with such content, on the basis of which he could issue perceptual judgments. The difficulty with this conversion is that there could be no question of the content arrived at being faithful to the original experience, because faithfulness is a relation which can only obtain between contents which are propositional in form. If a judgment is (isn't) faithful to an experience, it must be (un)faithful to how that experience is: so there must be some way that experience is; it must have conceptual structure. Echoing Wittgenstein, we can say that in the envisaged act of conceptualization whatever seemed right would be right. And that only means we can't talk about 'right'. Or, as Williams puts it (p.102), the given 'cannot provide a rational check on anything.'

What about perceptions 'as of'? These are certainly conceptual mental states, so they escape the above objections. But what reason have we to suppose that they specify adequately the contents of any mental states, even from the inside? We can direct our intuitions here by pressing the following question: as I write, am I enjoying a perception as of a pen or as of this pen? The friend of perceptions 'as of' must reply that I have a perception as of a pen, for a perception as of this pen just is a perception of this pen. If my perception is fully conceptual, and if this pen is its intentional object, what does the insertion of an 'as of' rider either add or subtract? The whole point of the 'as of' manoeuvre is to remove existence-dependence from the mental states it qualifies. It is to provide for the possibility of contingent propositions about

connections between items in minds and propositions about items in the world, and hence for the possibility of causal relations between items from those groups.

So my 'as of' perception is as of a pen. But is this an adequate characterization of my mental state? The suggestion is that the contents of my mind have a purely general characterization, that the de re connection can be hived off and located elsewhere - precisely in a causal connection between mind and world. This seems to me a fundamental falsification of what it is to perceive a particular. The view that the contents of mind can always be captured in purely general terms - a view beautifully expressed in Wittgenstein's remark that if God looked into our minds He would not be able to see what we were thinking of (PI II.xi) - is one that cannot survive proper attention to the nature of perceptual phenomenology.

We must make a distinction here between the capacity of thought in general to concern a particular and the capacity of perception to do so. The capacity of thought to do so is connected with the possibility of massive duplication (i.e. the possibility that there could be a universe exactly similar to our own): in the context of such a possibility, the availability of de re thought is the only thing which guarantees that I am thinking about a particular in my universe rather than a particular in the symmetrical universe. This is the topic of chapter 4. Here we are only interested in the capacity of perception to concern a particular; this capacity has to take account of the possibility not merely of massive duplication: partial or local duplication would be quite adequate to undermine

the capacity of perception to present subjects with objects, were the relations between perception and its objects construed along the lines favoured by the friend of perceptions 'as of'. That is to say: were perceptual mental states specifiable purely in general terms, my ability to perceive this pen I am holding would be undermined by the availability (in principle) of a qualitatively identical visual experience to someone else holding a qualitatively identical pen. We would have no principle of distinction between the objects of these two experiences.

Now such matching experiences obviously are available; and yet there is no doubt that it is my pen I am looking at. What settles which pen I am looking at cannot now be a matter of which pen is causally responsible for my experience construed as, in itself and from the inside, purely general in character: for it is simply not coherent to suppose that perceptual phenomenology might be merely general in content. What on earth would it be like to have an experience of some pen or other, without experiencing a particular pen - a pen which one could (at least mentally) pick out and point to (or at least seem to oneself to point to)? It is legitimate to ask after the 'inner' aspect of such general mental states, because they do essentially have phenomenologically characterizable 'inner' aspects: it is like something to be in such a state. (No one seriously denies that such aspects exist; even materialists only want to give an account of ^{what} they consist in.) Of course when I perceive this pen I also perceive a pen, and so could be said to have a perception as of a pen. But the point is that that cannot be the whole story about my mental state. Indeed it cannot be the whole

story about any perceptual mental state. For there is simply no way of perceiving some pen or other (having a perception as of a pen) without perceiving a particular pen demonstrable in the de re mode. Perceptions 'as of' are in the first instance perceptions de re. Again, it is no good arguing here that from within my perceptual experience the only pen I can demonstrate de re is an intentional, but possibly not real, object. If my state of seeming is a state of seeing, then my intentional pen is a real pen. There can be no question of there being two pens here. (I shall come the problem of hallucinatory states - states where the seeming is not also a seeing - and the associated sceptical possibilities shortly.)

As long as we are clear on this, we need not cavil if the friend of perceptions 'as of' seeks to locate a causal relation between a particular and the general component of a perception of it. There is indeed a contingent connection between this pen and my current mental state of perceiving some pen or other. And I have no objection to this relation being called causal. (This is actually all Grice wants in the article cited above.) Still, one might object that finding a causal relation here is pointless. The relation would be doing no explanatory work: it would not be embedded in any kind of theory. My point is just that a general characterization of such mental states is necessarily only a superficial characterization, and that when we give the complete picture causality must withdraw from it, to be replaced by the 'logical' relation of intentionality.¹⁰

It should be noted that I have not claimed that just any essential or necessary relation between two objects rules out a

causal relation. That would be absurd, for between any two causally related items there presumably obtains at least the essential or necessary relation of non-identity. Rather, the position has been this (at the risk of labouring I summarize). Suppose we start with a Humean notion of causation to this extent: we say that causal relata are logically distinct existences (the statement of the existence of the one relatum does not entail the statement of the existence of the other). This is a good notion of causation to work with, because it is the one natural science uses. Now it follows that no object or event can be such that there is another object or event which necessarily causes it.¹¹ If S perceives perceives A, it necessarily follows that A exists: so A cannot cause S's perception. Even if S's mental state is characterized as a state of seeming to perceive A, the fact remains that if S perceives A, S's state of seeming is a state of seeing, so that the 'effect' - characterize it how we will - essentially involves the 'cause'.

At this point the following objection is frequently lodged: surely it is precisely the fact that (among other things) A is causally related, in a suitable way, to S's mind that constitutes S's mental state as a perception of A. But this objection is misconceived. For what is the sense of 'causally related' here? It cannot be the Humean sense we mentioned above (and we are not here concerned with any other sense): if it were, it would be possible to specify the effect without bringing in the cause, and that is not possible. The only thing in S's mind which can plausibly be regarded as a candidate to be such an effect (because it is the only thing which is there in his mind) is his perception (or apparent

perception) of A: given that his apparent perception is a perception, each of these mental states commits us to the existence of A. If we tried to specify a mentation which really could figure as a Humean effect, we should be forced to ascend to the purely general, and arrive at something like: S's seeming to perceive (or perceiving) an F (assuming that A is an F). But that on its own is just not recognizable as the content of a mental state at all: no candidate to be a perceptual mental state could have a purely general content. Nothing in S's mind can count as an effect because anything locatable there already contains A as an intentional object. The only safe place to look for an effect of a causal transaction in our scenario is S's brain. I return to this in 1.5.

1.2 Error in Perception.

I shall now honour the promissory notes left in the previous discussion. If the presence or absence of a causal connection is not what makes for success or failure in perception, some account is owed of what success and failure do consist in. What, for example, has gone wrong when I suffer an hallucination? I shall tackle first the cases of hallucination and the so-called 'global' sceptical possibilities that I might be dreaming, or that I might be a brain in a vat on Alpha Centauri (being stimulated by a suitably ingenious experimenter to have the experiences which I have). Let us take the sceptical possibilities first.

It is often alleged that what rules out these possibilities is the absence of a causal connection between the purported mind and its purported intentional objects. Putnam argues, for example (chapter 1), that if I were a brain in a vat I would not be able to wonder whether I was a brain in a vat, since the thought I was essaying would not enjoy the kind of causal connections with brains and vats necessary to my being able to think about brains and vats at all. So if I can wonder whether I am a brain in a vat, the answer must be that I am not. And similarly for dreaming. The suggestion is that the sceptical possibilities are self-refuting.

At this point I would like to make the promised qualification to my acceptance of the Gricean point. What I conceded above was that there might be a causal explanation of why one enjoyed a

perception of one object rather than a (different) perception of another object. In the case of the above example, I agreed that we can explain why S has a perception of A but not of B by adducing the fact that his brain is causally related to the object correlated with A and not the object correlated with B. (I shall return to this relation of correlation in 1.5.) But the qualification that needs to be made to this is the following: we do not have to descend to the causal level to explain why S has the one mental state rather than the other. In the first instance, the reason why S perceives A rather than B is that A is in his (distorted) line of vision, whereas B is obscured by an opaque object. This is not a causal explanation: it is a logical truth that you can only see objects in your line of vision.¹²

This truth constitutes part of a simple theory of perception which we must ascribe to any subject of perceptual experience. The theory specifies what are enabling conditions of perception (see here Evans [1], pp. 88-9). Some of these conditions relate to the way the world must be if it is to be perceived: for example there must be adequate lighting, absence of distorting factors and interference. Other conditions relate to the state of the subject: he must be suitably attentive, facing in the right direction, and so on. The necessity of having such a theory emerges when we ask how the subject must think of the world if he is to enjoy a representational experience of it. He must think of the world as independent of his experience, in the sense that the very same kind of state of affairs which he sometimes perceives is capable of existing unperceived (even if it never actually does so). It would

appear that the only way to achieve this conception is to think of perceptual success as dependent (q.v. infra) on the fulfilment of the enabling conditions; so that one makes sense of the possibility of the existence of unperceived objects in terms of the thought that those conditions may not be met. That thought cannot in its turn be merely tantamount to the thought that one did not perceive the feature in question, because we would not then have the material to distinguish in thought between failing to perceive a feature which is there to be perceived, and failing to perceive a feature which is not there to be perceived.

But it is crucial to observe here that the dependence in question (of the success of perception on the fulfilment of the enabling conditions) is a logical and not a causal dependence: it is a logical truth that where perception occurs then to the extent that it does occur (there may be a continuum here) at least some of those conditions were met. We shall see below how appeal to the enabling conditions may help us to elucidate perceptual error, but at this stage all I want to insist on is the logical connection between perceptual success and the fulfilment of those conditions. This logical truth does indeed have a causal underpinning (in the Gricean case the behaviour of light waves), but that is an entirely adventitious feature of the account. We may compare the truth that no object can look red and green all over. Again, there is a causal underpinning of this truth, but it is in the first instance a truth yielded by the logic of colour concepts as they are fixed phenomenologically: it cannot be reduced to a natural truth (see McGinn [2], chapter 3). Nothing rules out the possibility that there

might, in certain conceivable situations, be no such causal underpinning to either of these conceptual truths. There would be nothing mysterious about such situations: there would be nothing mysterious about a world in which natural science as we have it could not be practised. An explanation of the Gricean scenario which stopped short of descending to the causal level would not be inadequate; it just so happens that in our world there is more to be said.

If this is right, it follows that Putnam must be wrong about brains in vats. A thinker can entertain thoughts about brains in vats even if he is in no causal contact with them, or indeed with anything at all. It follows that as far as these considerations go the sceptical possibilities are both genuine and coherent: that I am not causally linked up to anything in that world, if I am not, is of no consequence, because it would not be of any consequence if I were so linked up. The linkage, if it obtained, would not be between my world and my mind - that connection is 'logical' or essential - but between whatever my world supervened on and my brain. And the logical connection at best supervenes on a causal connection: it is not constituted by it, and so can exist in the absence of the causal underlay (dreaming), or in the presence of a non-standard causal underlay (brains in vats). It follows that we must account for what makes hallucinatory states hallucinatory elsewhere than in the absence of a causal connection.

First we must ask: how should we characterize hallucination? Here it is natural to say that when I hallucinate I have a visual experience, or at least seem visually to myself to have an

experience, of an object. In 1.1 I claimed that there is no way of perceiving a pen (some pen or other) without perceiving a particular pen demonstrable in the de re mode. I am committed therefore to ascribing genuine singular thoughts to a hallucinating subject, provided that his phenomenology is coherent enough to warrant such an ascription, and taking it that whether or not a subject enjoys a sufficiently rich phenomenology is a matter of his sincere say-so; for it could hardly be a matter of anyone else's say-so. In other words I am committed to advancing beyond the characterization 'It seems to S as though he is perceiving an F' to 'S perceives a seeming-F'; or from 'it seems to S as though he is perceiving an object in the de re mode' to 'S is perceiving an (intentional) object in the de re mode'.

What is the objection to this? That I am double-counting (i.e. counting the hallucinator's experience as an apparent perception and as a perception)?¹³ But what is the objection to double-counting? I do not lose grip here of the truism that for attempted thoughts in the de re mode, if there is no appropriate object there is no thought. For it all depends which thought we are talking about. If S hallucinates, we can say that there is an object which he sees, namely the sensory object. But he may well take the sensory object to be an object in public space, and in that case he will be mistaken. His thought 'That man is coming to get me', say, will not be a thought about a public object, because there is no such object; if S takes his de re thought to concern such an object, he errs. There is no object of that sort, so there is no de re thought about an object of that sort. But it would be a mistake to conclude from

this that there is no object of any sort, and hence no de re thought of any kind to be had by the hallucinator. It would be a mistake to say of the hallucinating subject that while

his mind is [not] wholly vacant; images and words may clearly pass through it,...[nevertheless] there is a kind of thought we sometimes have, typically expressed in the form 'This F is G', and we may aim to have a thought of this kind when, in virtue of the absence of any appropriate [i.e. physical] object, there is no such thought to be had...All that is being credited to such a subject is the intention of thinking a thought of a certain particular kind, and the belief that he is thinking such a thought. Obviously if there is no thought of the appropriate kind available, then there is no possibility, either for the subject or for anyone else, of giving the content of the thought he wishes, but fails, to entertain. (Evans [2], pp.45-6. Cf. McDowell [5].)

Now this line may have some plausibility in the case of non-perception-based thought about objects. Prior, against whom Evans is arguing in this passage, suggested ([2] p.154) that I could not think I had a de re thought about x without eo ipso having one. It may be an adequate response to this claim to deny to the thinker, in the case where there is no object, even the capacity to suppose that he means, or has a de re thought about, x. But the situation is quite different with thought about an object given, or apparently given, in perception. We cannot just credit a hallucinating subject with the intention to think a thought 'of a certain particular kind', i.e. a de re thought: for one thing, the subject may not possess the concept of de re thought. But the central point is that if we start with a subject who is really hallucinating, that is to say having or visually seeming to himself to have an experience of a pen (say) when there isn't really one, then it is not coherent to deny this subject genuine singular thoughts about the hallucinated

pen. We cannot allow him the seeming - the images passing through his mind - without allowing him the singular content of those seemings. Of course he cannot have de re thought about a physical pen; but he can, if the phenomenology is good enough, have a genuine singular thought about a visual (intentional) pen. The price of saying that the hallucinator cannot have de re thoughts at all is, after all, that we must say his mind is simply blank. But that is surely to take back the ascription to him of an hallucination. We do not, however, need to make this retreat, for once it is seen that the truism about singular thought is not being contravened, there should be no objection to double-counting. The mistake which it is easy to make with the defining principle of de re thought - 'no object, no thought' - is to try to take it substantially, to make it exert real leverage on actual cases. But it is essentially a bare principle. It cannot be used to favour any particular kind of object: it is ontologically neutral.

We are now in a position to tackle the question: wherein lies the error in a misperception? (Or: what makes an hallucinatory experience hallucinatory?) In the first place I think we should recognize a kind of error which is not intrinsic to the content of the experience itself but resides in the judgment the subject makes about the status of his experience. Plausibly, this is the sort of error I make when dreaming or would make if I were a brain in a vat. When I am dreaming I characteristically take it that I am not dreaming; if I were a brain in vat I would judge that I were not. There is nothing in the experience which warrants the judgment, and hence could be responsible for the error. Also within the category

of judgmental error, we can find error of the sort which the following episode illustrates. I am in a foreign country and I am acknowledged by someone whom I do not at the time recognize. It is only afterwards that I realize that the person was So-and-so, whom I know from my domestic haunts. Now in coming to this belated realization, the content of my experience need not change: I effect the recognition precisely by scrutinizing my memory image. My experience had a certain content which I failed to register at the level of judgment. Similarly, I may judge hastily that my experience has a certain content which later reflection on it reveals it not to have had. Another kind of case of judgmental error is merely verbal error (slips of the tongue, ignorance or forgetfulness of the meanings of words, etc.). This sort of mistake is corrected without impugning the experience on which the judgment is based.

Now it may look as if misperception cannot be fitted into the category of judgmental error. There are, for example, cases of illusion and hallucination, other than the radical hallucinations envisaged in the sceptical possibilities. Illusion occurs where the perception is faulty at the focus of attention: I glance momentarily at a piece of mud on my doorstep and mistake it for a letter. In cases of hallucination, on the other hand, the perception is good at the focus of attention, but faulty at the edges: I see the dagger before me all right - there is no mistake there - but I also experience it as existing in the same space as the other objects around me which are not currently the focus of my attention, but of which I am nevertheless experientially aware (the battlements, the flagstaff, the scudding clouds overhead). What account are we to

give of error apparently lying not in judgment about the perception but in the perception itself?

The question is particularly pressing for me because of the line I am defending on (1) the nature of the hallucinator's experience - i.e allowing him intentional objects of his experience, and (2) the connection between representational experience and the world. This connection is, I am arguing, 'logical': the world just is the content of representational experience. It might then be thought that the possibility of error could not be accommodated in the phenomenology itself in so far as it is representational. For then, one might think, the possibility of error in the representation qua representation would be the possibility of error in the world. And that is not coherent. But, after all, the hallucinator does on my showing enjoy representational experience: so where does he go wrong? -

To solve this difficulty we need to appeal to the enabling conditions of perception which we earlier saw had to be specified in a theory of perception available to any subject of experience. Now it may be a consequence of the failure of the enabling conditions fully to be met that certain features of the experience present themselves as salient, and that were those conditions adequately met, just those features would not be salient (perhaps they would not exist), and others would. Thus the highlighting of certain features and the shading of others in consequence of bad lighting, are directly responsible for my taking a piece of mud to be a letter. Of course I may not make a mistake in unfavourable conditions: the obtaining of all the enabling conditions is a

logically sufficient, but not necessary, condition for success in perception. (The obtaining of some enabling conditions is a logically necessary condition for perceptual success.) The enabling conditions for perception are in fact just those conditions which can be blamed for not obtaining when perceptual failure occurs. (There is no prospect of constructing an exhaustive list.) It is precisely this fact which supplies us with our account of perceptual error. For what it means to say that not all enabling conditions have been met is that the faulty perception can be corrected by perceptions which are not hampered by the failure of those conditions to obtain. Faulty perceptions are those perceptions which can be controverted by reliable perceptions. The inclusion of an account of the role of enabling conditions gives us a way of understanding this truism: a reliable perception controverts a faulty perception just when the enabling conditions are improved.

Now this account need not push us in the direction of a coherence theory. Error is located between experiences and not within a single experience, but that doesn't mean that all experiences are on a par. For some experiences - and this is the crucial point - are indefinitely corroborable, others not. The world, one might then say, is the content of indefinitely corroborable representational experience. This makes clear one sense in which the world - along with its linguistic correlate, truth - is an ideal. Our experience aims at matching this ideal, but the possibility of failure is always with us. Given that we must accord genuine singular thought even to an hallucinating subject, representational experience does at least present the subject with a a

world, but it can mislead the subject about the world: it can be 'one-off', or uncorroborable. In a sense, then, we do construe experiential error as a species of judgmental error. An hallucinator who is taken in by his hallucination makes an error of judgment: he judges that his experience is corroborable, when it is not. His experience does not, as such and in itself, lie. The experience cannot declare on the question of its own status; and it is there where the error can be located. Of course an hallucinator need not be taken in by his hallucination; but the error still does not reside in the experience, but rather in the hypothetical judgment - which the subject would make if he took the experience at face value - that it is corroborable.

The account of error I am offering is in one sense trivial, because it simply moves through analytical connections. But I claim that it is, even so, elucidatory. The challenge is to make sense of experiential error within the constraint set by the truth that representations qua representations cannot lie about their own content. That challenge is met by locating error in the relations between representations, and by identifying the world as the content of reliable perceptions: misperceptions are those that would be controverted by reliable perceptions; reliable perceptions are those which are indefinitely corroborable. Although this is just a truism, it is elucidatory when set in the context of the (again truistic) implication from fulfilment of the enabling conditions to success in perception. Enough analysis is explanation; we are not obliged to seek elucidation outside conceptual links.¹⁴

1.3 The Internalist View

There must be a sense in which the hallucinator inhabits a world of objects. That world is not the world, because the world is what an ideal perceiver (i.e. one who perceived only indefinitely corroborable experiences) would perceive. But we have seen reason why we should permit non-ideal perceivers to inhabit temporary or provisional worlds, fitted with temporary or provisional objects (the intentional objects of their mental states). We do not herein embrace ontological neutralism, but we do subscribe to a certain toleration. An hallucinator fails to engage with the world, but he does engage with a world, and a world which in some cases (those where the phenomenology is rich and coherent enough) might, for all the subject knew at the time, have turned out to be the world. He did not make a metaphysical mistake: he did not fail to find any world at all.

I have said that the principle constraining de re thought, 'no object, no thought', cannot be used to exert substantial leverage in actual cases. It is just a defining principle: there is a species of thought (genuine singular, or de re, thought) to which it applies. I have insisted that if we grant the hallucinator experience at all, we must permit him to entertain genuine singular thoughts about the objects of his experience. Of course it is true that he cannot, within the scope of the hallucination, entertain thoughts about the world. Nevertheless we do not want to say that, from the point of

view of the ideal perceiver, his mind is simply blank, on the grounds that there is, in the world, no object in the place where he is looking. Nor do we want to say that the content of the thought we allow him is purely general: to the effect that there is an object of such-and-such a sort positioned at such-and-such a location. No experience of any sort could merely have that content. We want to ascribe to the hallucinator a thought which is close enough to the thought essayed to represent faithfully his experience and intention, but which falls short of the very thought essayed, that being unavailable to him. (I assume here that the hallucinator, in ignorance of the fact that he is hallucinating, is essaying a thought outside the scope of the hallucination: that is the interesting and problematic case.)

It seems to me that we can capture the facts of the above scenario in terms of a distinction, which I have already sketched, between provisional and ideal worlds. The hallucinator essays a genuine singular thought: within his provisional world he succeeds; he entertains a thought about (as he would put it) that elephant. Of course the relativization of his thought to a provisional world is indispensable here: he cannot entertain a singular thought about an elephant simpliciter, because there is no elephant. But he must be having a de re thought about something: to construe him as having in mind a general thought about some unique elephant is to stray to far from his actual and acknowledged experience; it is radically to misconstrue his cast of mind. (Apart from anything else, this will come out in the way he reacts on learning of his mistake: erroneously supposing that there is some unique elephant somewhere

is considerably less disconcerting than thinking, wrongly, that you see one.)

To accord the hallucinator representational experience is to give him a world of a kind. In dubbing him, rather than us, 'the hallucinator' we are putting our money on the veridicality of our experience (the survival-prospects of our provisional world) as against his; we might of course be wrong about this. The position I am taking up has phenomenalist leanings, since it tolerates a certain equality between all representational experiences, and it regards the world as the content of (a certain subset of) representational experiences; but it is also realistic: the realism is registered in a rider to the effect that only some - privileged - representational experiences count as yielding the world as their content, namely those which are indefinitely corroborable.

We could say that on this picture the subject of experience is confronted, in the first instance, with intentional objects, or appearances with objective content; some of those intentional objects may be real objects - some of the appearances may be corroborable - some not. To have a veridical experience - to be hooked on to the world - is to be confronted with intentional objects which are also real; or, we might say, it is to be confronted with appearances (or have experiences) which are indefinitely corroborable. (These ways of putting it are just variants of one another. In particular, to talk of confronting an appearance with objective content is, on this view, just to talk of having an experience with an intentional object - the question of whether the subject is so confronted being settled by the nature of

the phenomenology of the experience itself.) This picture contrasts with one according to which the mind is conceived to be either in direct touch with the world (when the experience is veridical), or only apparently in touch with a world, and not in touch with anything else (when the experience is hallucinatory). On this alternative picture, the hallucinator is not conceived as inhabiting a world, not even a provisional one: he is in a state of its sensibly seeming to him that he inhabits a world (that he is in a position to entertain de re thoughts). That the alternative picture misrepresents the nature of states of seeming, I have already made clear. There are not, I think, any compensating features of the alternative account. It is no help, for example, in dealing with scepticism. On that front each picture meets with the same level of success (or failure). Whether we conceive of the mind as, in the first instance, confronting appearances (and then the world is the content of ideal appearances in the explained sense), or whether we conceive of the mind as confronting either the world (if the experience is veridical) or merely appearing to confront a world (if the subject is illuding or hallucinating), is a matter of indifference to the issue of what the world is, or how much we can know.

Metaphysically, on either picture the world must be acknowledged to be essentially capable of being experienced. The first - internalist or first-personal - picture accords a certain authority to the individual subject: it allows the test for representationality in experience to be applied by the subject himself; he can tell from the inside whether or not his experience

is representational. (This is of a piece with according the hallucinator representational experience, since no one other than the hallucinator could have authority on the content of his experience.) And he can therefore entertain thoughts about its intentional objects. The second - externalist or third-personal - picture ties representationality in experience to perceptual success: only the veridical perceiver has representational experiences and thoughts about the objects of those experiences (all others merely seem to have them). And of course perceptual success is a matter over which the subject has no special authority. Epistemologically, according to the first picture, the subject is acquainted with the objects of his representational experiences - the possession of which latter he is the principal judge. He knows that his intentional objects are thus and so; but he does not necessarily know (although he can do so up to scepticism) what their status is (because he doesn't necessarily know what the status of his representational experiences is). So he does not necessarily have knowledge about objects in the world, only about objects in his provisional world. According to the second picture, the subject's acquaintance with objects of experience - his ability to know that they are thus and so - is supposed not to be a matter merely of what is internal to his experience, but also partly a matter of what is there in the world. Up to scepticism he can know about these objects. But if he fails to have such knowledge, he does not, as a consolation prize, succeed at least in knowing about a world of purely experiential objects. He seemed to be acquainted with real objects, but was not.

The point of inserting the proviso 'up to scepticism' in the above discussion, was this: on each picture, a subject may achieve a position where his experiences are rich, coherent and continuously corroborated by others. This is the position of most of us most of the time. We occasionally have rogue experiences, but generally they are easily and quickly isolated, and then discounted. When a subject reaches this position, only a global scepticism of the sort mentioned in the last section can dislodge him. Failing such a scepticism, this subject has knowledge of his environment, on the first picture because his experiences will, as a matter of fact, be endlessly corroborated; on the second picture because the de facto failure of the scepticism means that he must be experiencing the world, not merely seeming to do so. And on the other hand, as long as global scepticism is (perhaps judged unlikely but) unrefuted - the only actual position we can envisage - then the most we can say is that, on the first picture, the subject has knowledge of his provisional environment, but does not know whether his experiences of that provisional environment will be endlessly corroborated (he may wake up from the dream); and on the second picture the subject seems to have knowledge of the world, but this purported knowledge may (if he wakes up) turn out to be bogus. I cannot see a significant epistemological difference between these two positions.

In the meantime, the subject must work with what he has: on the first picture, he has (possibly corroborable) knowledge of a provisional world; on the second picture, he has the (possibly veridical) appearance of knowledge of the world. On the first picture, our working epistemic notion is knowledge of a

(provisional) world, and the ideal towards which we aim is knowledge of the (real) world; on the second picture, our working epistemic notion is the right to claim knowledge (on the basis of appearances), and the ideal towards which we aim is knowledge of the world (the only world there is). Again, as far as epistemological considerations go, these positions look like mere notational variants of one another. The debate between internalists and externalist does not have the metaphysical or epistemological bite that has been ascribed to it.¹⁵

The discussion so far is reminiscent of another and related debate in the philosophy of mind - that between criteriologists and disjunctivists - and the irenic approach I have adopted towards the 'conflict' between internalists and externalists is equally available in this case. As applied to representational experience, the debate looks like this: Criteriologists are internalists: they picture subjects of experience as confronted, in the first instance, with appearances to the effect that the world is thus and so. The appearances are constitutively related to the world (they have conceptual content), but they are not self-certifying in respect of their veridicality. In other words, they are defeasible criteria for the obtaining of the state of affairs which they represent. So, if I look out of my window and seem to see rain, I am confronted with an appearance precisely of rain (the special, criterial connection between appearances and the world); but the criterion may be defeated by being embedded in suitably uncorroborative further conditions. The slightly unorthodox sense of the word 'criterion' employed here derives at least from discussions of Wittgenstein's

use of the term, and perhaps from that use itself.¹⁶

Disjunctivists argue that my experience in the above case is constituted either by my really seeing rain, or by my merely seeming to see rain. This disjunction is fundamental to the experience: there is no simpler characterization of that experience. Note that there is not supposed to be any implication to the effect that I can never know which disjunct I am located on; of course I often can (although only up to scepticism). The aim of the disjunctivist manoeuvre is to put the world back into the cognitive reach of subjects: phenomenologists are supposed to divorce the subject from the world, trapping him behind (mere) appearance.¹⁷ But it should be clear that these positions are for metaphysical and epistemological purposes equivalent variants of one another. The criterialist's subject is not divorced from the world, as his opponent claims, precisely because of the special, internal connection between the appearances he faces and the world. If those appearances are undefeated after subjection to reasonable testing, then they count as veridical - only up to scepticism, it goes without saying, but then the disjunctivist can get no further in his confidence (after lookalike processes of examination) that he is located on the disjunct of veridicality rather than the disjunct of illusion. To accuse the criterialist of idealism is just not to take seriously the special sense of 'criterion' he is operating with.¹⁸ The criterialist is, plainly, the internalist of my earlier account: so far undefeated criteria are so far corroborated appearances. And the disjunctivist is just the externalist. The version of phenomenology espoused by the internalist camp - which we might call

'logical phenomenalism'¹⁹ - does not conflict with realism, but coincides with it.

'Logical' phenomenalism is just the doctrine that the world is the content of the logical combination of appearances. We must add that the world is the content of the logical combination of corroborable appearances (the concession to realism). Of course the two accounts of what it is for, say, there to be a table in the next room do differ slightly in presentation. The realist starts his picture with the bare existence of the object: that is a segment of the world, and is what is cognitively given to ^a subject of experience. The phenomenalist starts his picture with the content of actual and possible table-experiences (the intentional tables) available to subjects, if and when they go into the next room; he then works over to the existence of the object, which he specifies in terms of the indefinite number of corroborative experiences available to (ideal) perceivers. But although these positions start differently, they converge under supplementation. The realist can hardly deny that the existence of the object is partly constituted by the infinity of actual and possible perceptions of it which its existence makes available. These perceptions are a clear necessary condition of its presence in the next room. For his part, the phenomenalist can hardly deny that what all the perceptions on his infinite list have in common (what collects the list) is their common possession of that table as their intentional object. Metaphysically and epistemologically, first-person and third-person perspectives (internalism and externalism) point at one another.

The only (but paramount) respect in which the externalist's

position is inferior to the internalist's is in its misrepresentation of what it is like to have a visual experience (what it is like to seem to see something); in denying the availability of de re thought to someone who merely seems to see something (but who still does seem to see it), the externalist is giving with one hand and taking away with the other; for no such experience is phenomenologically possible. This misrepresentation is not a light matter. To correct it, an externalist could renounce externalism and embrace internalism. But he has another option. He could adopt a more radical externalism, refusing to allow that a subject who fails to see a real object, may even seem to see anything. After a brief discussion of another matter, I shall return to this sort of externalism in the final section of this chapter.

1.4 Appearances

My claim has been that what we perceive are, in a certain sense, appearances with objective content. I have endorsed the phenomenalist's or internalist's picture of perceptual knowledge over the externalist's, because the externalist misrepresents the phenomenology of experience. Now to say that the subject is confronted with an appearance of an object is just to say that he has an experience of that object. The object is not itself the appearance: it is the content of the appearance. On the internalist's view, the subject can have such an experience - he can be presented with an appearance of an object - even if he is hallucinating. The point of introducing the 'appearance' terminology - instead of continuing to talk exclusively in terms of experiences and their intentional objects - is to stress that there is a phenomenological similarity, for the subject, between an experience which is hallucinatory and one which is veridical. They are or can be indistinguishable by him.

It is important to observe that on the phenomenalist's picture it would be wrong to suppose that the subject is confronted with appearances of objects which are themselves construed as lying behind the appearances: rather, he faces appearances with objective content. (Here I diverge from Kant: see the preface to B, p.xxvi.²⁰)
When I look at my pen it is an appearance of that pen I see, not a

mere appearance of a pen (with the real pen lurking unreachably behind, as it were) - the very same pen which I feel when I hold it. If a spatial metaphor is felt to be necessary, it would be more appropriate to say that an experienced object is in the appearance, rather than behind it. Appearances do not mediate between objects and minds, in the sense of getting in the way: they figure in the model not only to register the continuity, from the inside, between hallucination and veridical perception, but also to stress that an object of experience, such as my pen, is a manifest object, that is: an object essentially capable of entering the content of representational experience. (This is in line with Kant's use of the terminology: see B.69-71.)

The terminology of appearance has historically been misleading, because to say that what the mind confronts in perception is an appearance is suggestive of an attenuated degree of reality attaching to the objects of those appearances. And although the charge which the disjunctivist levelled against his opponent - that he cut the subject off from the world and left him trapped behind mere appearances - was, as we saw, a piece of mislocated rhetoric when applied to the criteriologist or the 'logical' phenomenalist; nevertheless the charge does stick to some versions of the sense-datum theory, such, for example, as one encounters in the opening pages of Russell's essay The Problems of Philosophy. The case is interesting, and I want to use it to help dissociate my employment of the 'appearance' terminology not only from any suggestion of attenuated reality, but also from two other supposed implications of a phenomenalist view - that objects are perceived

indirectly, and that they are private to the perceiver. All three of these implications intrude into Russell's theory of knowledge.

Russell starts with the platitudinous enough thought that an object, such as a table, characteristically presents different appearances to different observers differently placed:

If several people are looking at the table at the same moment, no two of them will see exactly the same distribution of colours, because no two can see it from exactly the same point of view, and any change in the point of view makes some change in the way the light is reflected. (p.2)

The point is not confined to the sense of sight, but has quite general application to perception (Russell himself gives further examples relating to the sense of touch). Where other senses are in question, the phrase 'point of view' will have metaphorical application, but with this rider noted, we can obviously generalize Russell's claim so that it takes the following form: perceptual experience varies as point of view, understood broadly, varies. Russell draws three conclusions from this thesis. Firstly, because the table presents different appearances to different points of view, none of the appearances can present the real table. For example, in the case of shape, 'what we see is constantly changing in shape as we move about the room; so that here . . . the senses seem not to give us the truth about the table itself, but only the appearance of the table.'²¹

It has frequently been objected, against the sort of view Russell advances here, that although there is a sense in which a table viewed from different angles presents different quadrilateral shapes to the perceiver (the sense in which it presents different

shapes to the painter); nonetheless the table still looks rectangular to the perceiver from all these positions (given a suitable viewing environment), in the sense that rectangular is the shape he would, from each point of view, judge the table to be, given the way his visual experience is and on the basis of that experience. There is no question of inference from what is really seen; the table is, in this sense, really seen as rectangular in shape. In fact, to see the table as anything other than rectangular is something of a sophisticated achievement.²²

Now this objection is ultimately fatal to Russell's philosophy of mind here. But it does not take effect immediately. For the Russellian sense-datum theorist can simply adapt his claim to accommodate the objection. Even although it is conceded that the shape which a perceiver strictly and literally seems to see, in all the different positions from which he views the table, is a rectangle; still it must be allowed that the mode of presentation of that shape varies from point of view to point of view. By the 'mode of presentation' of the rectangle we mean, as I put it above, what the painter sees (or, more accurately, the shape which the painter must put on the canvass). Now the sense-datum theorist's point can be restated thus: 'none of the modes of presentation of the table is privileged, and they all differ from one another in their visual content. So none of them can present the real table, because there would be no reason to accord this honour to one of them rather than another.'

But now that we have restated the sense-datum theorist's point to cater for the above objection, we can see the mistake lying on

the surface. For of course all the variously angled perceptions of the table give us the real shape of the table. The table does look different from different points of view (it has different modes of presentation). But it is exactly the shape of the table which looks different from different points of view, and not the different (mere) appearances had from those points of view. One might say that it is what is common to all the separate appearances relativised to their points of view. Hence we see that the above move by the objector is really crucial, since it establishes, against our sense-datum theorist, that there is something common, in respect of shape, to all the various appearances. They are all appearances, in their various ways, of a rectangular table. There is no difficulty in attaching the 'real' to this common element, and so denying that there is here any opposition between appearance and reality.

Russell's second conclusion from his initial thought is related to the first: it is that the various perceptions of an object, had from different points of view, cannot give us direct epistemic contact with that object, but are 'at most...signs of some property which perhaps causes all the sensations, but is not actually apparent in any of them.' (op. cit., p.3). The causal theory is a natural concomitant of the radical appearance/reality divide which he has embraced. If one regards the objects of perception as appearances in Russell's sense (i.e. intermediaries which get in the way of direct cognitive access to the world), one seems forced to construe reality as that which is causally responsible for those appearances, but which is itself incapable of being experienced. This characterization fits the scientific world, but that cannot be the

only world we inhabit. For to suppose that the manifest world is not really a world is to lose one's grip on what reality, or a world, is. On the other hand, to hypothesize a non-scientific would-be manifest world behind the appearances, a world of tables and chairs which we cannot see, but which are causally responsible for the table- and chair-appearances which we enjoy, is simply incoherent.²³ Tables and chairs are essentially capable of being experienced. These considerations give us additional reason to reject the causal theory of perception, since there seems to be a mutual implication between a causal theory and a radical appearance/reality split along one of the two untenable lines mentioned above.

Russell's third conclusion is that the appearances had by perceivers enjoying different points of view are private to the perceivers:

When ten people are sitting round a dinner-table, it seems preposterous to maintain that they are not seeing the same table-cloth, the same knives and forks and spoons and glasses. But the sense-data are private to each separate person. (p.9)

In one sense this point is correct. The perceivers have differently angled perceptions. Perhaps that is all Russell meant; in that case, the point is tantamount to the first one: from different angles, the table presents slightly different appearances. But there is also a suggestion in Russell's text that he meant 'private' in the sense which Wittgenstein was later to make famous: i.e. logically unavailable to anyone else. On this reading, sense-data would not be public, relatively enduring features of the world available to be picked up by anyone at all suitably placed, but would be fleeting, internal phenomena, whose identity-conditions would be such as to

make it necessarily the case that no two people could ever share one sense-datum, and one person could not have the same sense-datum after a temporal gap, but at best an exactly similar sense-datum. This conception of what sense-data must be has been very powerful among philosophers. But if it stems from the sort of consideration adduced by Russell in the above passage, then it is rather surprising that it should have achieved dominance. For one only needs to envisage the scenario in which Russell's ten people sitting round a table all move one place to the left, to scotch the suggestion that appearances have to be thought of as private. For in this scenario, the overwhelmingly natural thing to say is that each of the diners moves into the point of view, and hence is presented with the appearance, previously enjoyed by the person immediately to his left.

Of course some of the diners may differ visually from others, so that some (let us say) see a slightly different shade of brown from others. But these possibilities are purely contingent features of an individual's perceptual make-up; they introduce no element of logical privacy. (There is, for instance, no logical bar to surgical operations being performed on some of the diners' brains so that they are all enabled to see exactly the same shade of brown.) Given the contingency of this sort of perceptual difference between one observer and another, we make no relevant change in our scenario if we just assume that all the diners have exactly similar perceptual equipments. If we then label one of the positions at the table P, the obvious and surely correct thing to say about our diners is that exactly the same appearance is available at P to all the diners, and

whenever any of them sits at P, he enjoys that appearance. Of course we can, if we like, use the notion of appearance in a different way, individuating appearances by perceiver, and then we shall have to say merely that the diners enjoy numerically different but qualitatively identical appearances from position P. But this way with the terminology is not mandatory; and given the availability of the notion of a common appearance, we should repudiate the suggestion that to talk of appearances as the objects of perception is inevitably to import logical privacy into one's theory of knowledge.

It might be countered that my talk of objects of experience being, in the first instance, intentional objects, does commit me to just this privacy; for surely intentional objects are proprietary to the particular experiences in which they figure? And of course experiences are individuated by subject of experience (that is true). If intentional objects are individuated by perceiver, do we not then have a version of the privacy which renders the possession by perceivers of a common objective world unattainable? But the point is that, as I have indicated, in the case of a genuine perception, there are not two objects: the intentional object and the real object. In that case the intentional object is the real object. A subject of representational experience - including an hallucinating subject - is presented in the first instance with an intentional object. If his experience is veridical, that object will be not merely intentional, but also real. But in that case it will not be individuated by any particular experience in which it figured, and hence by a particular subject. A representational

experience presents its object as available, in the public world, to be experienced by anyone at all: it presents it as real. That is to say: subjects of representational experience who take their experience at face value will assume that the objects of those experiences are real. All such objects are of course intentional; some of them are, in addition, real; others of them are merely intentional. If an object is merely intentional, it will be individuated by the experience in which it figured; if it is real, it will be individuated by the set of actual and possible experiences (in principle an indefinite number) in which it does and can figure. A subject may not be able to tell from the inside which sort of object he is confronting: as far as he is concerned, his intentional objects are at best provisionally judged to be real. Or we might say that the appearances he confronts are at best inchoately objective: they are fitted up - they are suitable candidates - for objective status. He acts on the assumption that they are objective: if he is right, no question of privacy arises.

In general, the danger lurking in the use of the 'appearance' terminology is this (cf. note 19): if we allow the hallucinating or perceptually illuded subject representational experience with intentional objects (=if we allow him to confront appearances of objects), then, given that there is a phenomenological continuity for the subject from hallucinated objects (hallucinatory appearances) to real objects (veridical appearances), it may look as if we are forced to accord the same degree of reality (veridicality) to both sorts of object (appearance). Then the hallucinated object will be as real as the real object; the real object will be as

illusory as the hallucinated object. But these are empty terrors. On the version of phenomenalism which I have offered, reality of object (veridicality in appearance) consists in the indefinite corroborability of experiences, or appearances, of that object. So if a subject is presented with an appearance which is in fact indefinitely corroborable, he does not confront a 'mere' appearance, behind which he is 'trapped'; he confronts a real object in appearance. Two of Russell's characteristics - attenuated reality and privacy - do attach to objects of hallucinatory or illusory experience. But the fact that the subject may not be able to tell from the inside which sort of experience he has does not go to show that if his experience is veridical, still the objects of that experience are both unreal and private to him. On the contrary, in that case they will be fully real, public objects. This is the force of the concession to realism.

Internalism defeats externalism over the characterization of a subject's phenomenology; it is then modified in a realistic direction by incorporating ideality into the notion of the experienced world. Precisely because it incorporates this ideality - and so removes the question of the veridicality of the subject's experience from the subject's own instantaneous say-so - we free the subject from being 'trapped' behind his experiences. So long as we are clear on the dialectic here, we can keep the terminology of 'appearance', misleading though it might otherwise be.

1.5 Externalism and the Manifest Image.

The connection between the scientific object (the object causally responsible for the brain-state of a perceiving subject) and the manifest object (the intentional object of the perceiver's experience, as the internalist construes that experience) is, as I have claimed, one of correlation; we might also call it supervenience (since the properties of the manifest object supervene on the properties of the scientific object). I want first of all in this section to say something about why this relation is so deeply problematical. The problem is, generally, how a relation between two items one of which is essentially involved in the other (a perception of a manifest object) can supervene on a causal transaction (the relation between the scientific object and the brain state underlying a perception). Descartes in effect raises this problem in his Sixth Meditation when he observes that a perceptual mental state only emerges from a brain state, and not from the whole causal process terminating in the brain state. It is this fact which the hypothesis that I am a brain in a vat exploits. But the possibility envisaged by Descartes is even more worrying than this: he considers the brain state which is standardly caused by damage to my foot of a certain sort and which standardly gives rise to a mental state whose intentional object is a certain kind of pain in the foot. He observes:

It is true that God could have constituted the nature of man in such a way that this same motion in the brain would have conveyed something quite different to the mind; for example, it might have produced consciousness of itself either in so far as it is in the brain, or as it is in the foot, or as it is in some other place between the foot and the brain, or it might finally have produced consciousness of anything else whatsoever. (HR p.197.)

The possibility seems to arise of 'alternative' races who suffer from such mismatch. When injured in the foot their neurophysiological causal chains are so wired up that they enjoy quite irrelevant experiences with possibly quite irrelevant intentional objects. Perhaps such races existed and failed to compete successfully for survival. To block this suggestion (surely intolerable), Descartes relies on the goodness of God: He created a race (us) such that this kind of mismatch rarely occurs; He so disposes the causal chains that they originate in the physical states of objects suitably related to (crucially, in the same place as) the manifest objects perceived by the mind. If something injurious is going on in my foot, I feel something there rather than elsewhere, and equally importantly I feel a pain rather than a pleasant warmth.

Now given the supervenience thesis, something like this account must be right, extraordinarily enough. But that should make us hostile to the supervenience thesis and its associated internalism - in fact push us towards some sort of identity thesis. For it is incredible that the 'mismatch' possibilities not created by God should be possibilities at all. The nature of representational experience, about which I have taken an internalist line, and the absence of a causal relation between perception and

its objects, push us towards supervenience (as well as other considerations relating to the mental)²⁴ but it is an incredible thesis.²⁵ There seems to be a deep tension here between the idea of a manifest world as experienced by subjects and the idea of a world as investigated by science.

The supervenience thesis we are committed to is, I have said, an incredible thesis; but let us be clear how radically destabilizing of our intuitions the alternative is. Ruth Millikan has written of a version of this alternative:

...We would maintain that...every kind of awareness of is in part an external relation, the inside of the awareness - the feeling part - giving no absolute guarantee that it is the inside of a genuine awareness of relation. Consciousness, that is, does not contain within it or directly before it any objects of consciousness...There is nothing diaphanous about consciousness. An unsettling possibility! One implication would be that we are no more in a position to know merely by Cartesian reflection that we are truly thinking i.e that we or our thoughts intend anything, than that we are thinking truly. Absolutely nothing is guaranteed directly from within an act of consciousness. (p.91-2)

But this passage does not get the matter quite right, and it is instructive to see why not.

If we allow the subject authority on his seemings, that is to say authority on the question of how things seem to him to be, then it is not open to us to deny him intentional objects of those states of seeming. That is because it could not seem to a subject that an F (or the F) is G without its seeming to him that, as he would put it, this or that F is G. A subject could not enjoy a purely general state of its seeming to him that things were thus and so: it would always be possible for him to advance to a de re thought about the object of his mental state. This fact serves as a general constraint

on the nature of representational experience, and is one on which I have already laid great stress. The upshot is that if, like Millikan, you wish to deny the subject any kind of authority on the existence of the objects of his conscious states, you should not so much as accord him authority on his seemings (let alone his seeings), i.e. on whether or not it seems to him that...

But, it might be objected, why not allow the subject authority on his seemings, and hence on the existence of the intentional objects of those states of seeming, without allowing him authority on whether or not his seemings amount to seeings, and so on whether or not the intentional objects of his seemings are real objects? In the above quotation, some of Millikan's phrases appear to hint at this less radical line: we can perhaps allow the subject authority on what he is in some weak sense aware of (merely his intentional objects), but not on what he is genuinely aware of. (This is suggested by her talk of 'the inside of the awareness - the feeling part'.) Perhaps there is room for an intermediate position which combines internalism and externalism in this way: it allows the subject to have authority over how things seem to him to be, and hence, as I have argued, de re thoughts about the objects of those states of seeming; in the case of his seemings the subject will obviously also have authority over the individuation, or singling out, of the intentional objects of those states of seeming, since you could not be the one who said that or how things seemed to you to be, without being the one who singled out which things seemed to you to be that way. Where the subject's seeming is also a seeing, the combined view takes an externalist line, individuating the

perceptual object by causal origin of the subject's brain-state; of course the subject has no special authority on this matter. If this combined position were coherent, Descartes' problem would indeed be circumvented, since the objects of perception (logically) could not fail to coincide with the objects causally responsible for the brain-states which partially constitute those perceptions. The supervenience thesis, which precipitates Descartes' problem, would be supplanted by an identity thesis.

It is important to see that this less radical line (reminiscent of - indeed a materialist version of - the disjunctivism in 1.3, corrected to admit intentional objects of states of seeming) occupies an unstable position. We have agreed that if we allow the subject authority on his seemings, we must allow him to have de re thoughts about the intentional objects of those seemings. Now some of those seemings will be seeings, in which case their objects will be real. Of course, since the subject is not an authority on whether or not his seemings are seeings, he is not an authority on whether the intentional objects of his mental states are real or not. But some of them will be real. It follows that in those cases where they are real, the subject will have entertained thoughts, in the de re mode, about real objects. So far so good. At this point, however, the combined view generates a crisis of individuation. For it allows the subject authority over the individuation of the objects of his seemings but not of his seeings. But when the seemings are seeings, it will be a corollary of the combined view that the subject will both have and fail to have authority on the individuation of the objects of his seemings/seeings, given that it is not open to us to

say (what would surely be intolerable) that in such cases there are two perceptual objects - an intentional and a real one; and given also that it would be incoherent to accord the subject authority over the contents of his seemings only when those seemings are not also seeings. So the intermediate view collapses.

If now we recoil in an internalist direction, and allow the subject authority over the individuation of the objects of both his seemings and his seeings (although still not authority, of course, over whether his seemings are seeings), Descartes' problem will recur: it will be conceivable that the manifest events could have emerged from any old causal transaction - even a wholly inappropriate one - at the fundamental level. So in seeking an alternative to the theory of mind which generates Descartes' problem, we must embrace an externalism much more thoroughgoing than the combined view: we must refuse to accord the subject authority on anything, even on how things seem to him to be. And since one cannot, in general, be conscious without being conscious of how things seem to one to be, and so cannot be an authority on whether or not one is conscious without being an authority on how things seem to one to be, it follows by contraposition that on this radical externalism the subject is not even an authority on whether or not he is conscious.

But now what stands in the way of this radically externalist line? Evidently, our strong intuition that the subject is an authority on whether or not he is conscious. The shape of the dialectic is this: if we take an internalist line on the subject's authority over how things seem to him to be - and hence, as I have

argued, over the individuation of the objects of his states of seeming and of seeing - then, given the anti-causal thesis argued for in 1.1., it follows that Descartes' problem will arise in any world, such as on this view ours is, in which the relation of intentionality supervenes on a causal relation between a scientifically characterized object and a brain. If, on the other hand, you reject internalism, it is open to you to bypass the anti-causal thesis (you cannot deny it, but you don't have to), and to say that when S perceives A there is nothing but an external (obviously causal) relation between the scientific object and S's brain. That relation, you might say, is what S's perception of A consists in. The anti-causal thesis is not herein denied, because it is not envisaged that the effect in the causal transaction is S's perception of A - that is what the whole transaction is - but rather whatever brain events occur as end-products of the transaction. And there is no essential connection between those brain-events and the scientific object which causes them.

There is on this externalist account no conflict between intentionality and causality; but that is because intentionality as I have set up that notion (the relation between perception and its objects) is simply bypassed by the externalist's causal relation: the termini of his causal relation do not correlate with the termini of the intentional relation (as they do on the internalist's picture, thus generating Descartes' problem). On the radically externalist view we are considering, nothing in S's mind constitutes the effect: the mental event of S's perceiving A is just the whole physical transaction beginning with light hitting the scientific

object and terminating in the brain—events produced by that light entering the eyes. That the relation between perception and its objects is non-natural is of no more concern to this externalist than that any 'logical' or essential relation is non-natural. And now Descartes' problem does not arise because perceptual objects are individuated by causal origin of the subject's brain-state, and hence it is not conceivable that perceptions should have variable causal provenances. (So the radical externalist takes over the externalism of the combined view while jettisoning its internalism.)

For such an externalist, the position would then have to be not that you can be mistaken about whether or not you are conscious, in the sense that you can think you are conscious without being conscious. That would be flatly self-contradictory. Rather, the position would be that you have no authority on any questions relating to your own consciousness — not even on the question of whether you think you are conscious. Millikan should not say that 'we cannot know a priori or with Cartesian certainty that in seeming to think or talk about something we are thinking or talking about — anything at all' (p.10); that formulation just hands the internalist victory on a plate, for the reasons which I have run through. Rather, naturalists should say that we cannot know with Cartesian certainty so much as that we seem to think or talk about something. This is a very radical claim, which is indeed unsettling. Could I really be wrong — in the above sense — about its now seeming to me that I see a piece of paper? Could I be wrong in my belief that I am conscious? That is: could I really, at this moment, both fail to be

conscious, and so also fail to have the belief that I am (and the belief that I have that belief, etc.)? Remember that the claim is not the contradictory: whenever you think you are conscious, you might not really be conscious. The claim cannot be put in this form. The best way to state the claim is this: a subject is not an authority on whether or not he is conscious; he is not the one who settles that issue. Indeed the externalist has to go as far as saying that he has no say in the matter at all, because for the subject to have any say on whether or not he is conscious would be for him to have the final say on the matter. But if we can state the doctrine without paradox it is nevertheless very hard to believe.

It is also a corollary of radical externalism that we have to say (as Putnam wanted) that the hallucinator's mind, and the mind of a brain in a vat (if it has one) is simply blank. For otherwise an internalist line with respect to their states of seeming would be forced on us: if the hallucinator is allowed to ~~seem~~ seem to see something, who else could have authority over what he seemed to see if not the hallucinator himself? And then the above crisis of individuation would ensue, via the internal continuity of seeming from the hallucinatory to the veridical case. The externalist must then say not only that the subject has no authority on whether or not he is conscious, but also that it is not possible to have an experience which falls short of veridical perception of some sort: it is possible to misperceive, of course, in the sense of having a perception which is in some ways veridical and in some ways not; but it is not possible to be in a state of total hallucination; or, at least, it is not possible to have hallucinatory experience. That

conclusion certainly does nothing to enhance the appeal of radical externalism.

We may fruitfully compare here a possibility broached by David Pears in his discussion ([1], p.56ff) of the question whether objects of amusement are or can be causes of amusement. He notes that a Humean notion of causality would require the corrigibility of the subject's estimate of what amused him. This is because Humean causality requires regularity, and there is no guarantee in advance that the regularities will show in favour of the subject's own judgments about what amused him. Now that a subject might be wrong about the object of his amused reaction runs counter to our intuitions about this sort of occurrent mental state. Pears tries to resolve the problem by supposing that these intuitions of ours are not immutable. Perhaps we might come to regard amusement in something of the way we regard depression, where we allow that a man may mistake the object (which really is here the cause) of his depressive reaction, and that the object/cause of depression may relate to depressive states in something like a Humean way. I am not myself sure whether our concept of amusement could make this shift while remaining recognizably a concept of amusement, but I do not want to contest that issue here. Rather, I want to draw attention to the fact that no such shift is possible in the case of concepts like 'what the subject thinks amuses him' or 'what the subject thinks depresses him'. If he has sincerely-held opinions on such matters at all, the subject simply cannot be wrong about what he thinks amuses or depresses him. The only alternative to his being right about these things is his not even being conscious, and so his not being a

subject (at the relevant moment) at all. Here we register a crucial difference between being depressed (or, if Pears is right, amused) and thinking that x is what depresses (amuses) you. The former state may be corrigible; the latter surely is incorrigible. The situation then constitutes a specific example of what we must say in general about being conscious and thinking that you are conscious. The externalist has now to say that the above alternative is always a genuine one: it is always possible that one is not conscious.

The internalist claims that this possibility sometimes fails to obtain: for example, that it is not now possible that I might, at this very moment, be unconscious. The fact that the internalist's claim can only be formulated using indexicals such as 'now' - it becomes false if a date is substituted for the occurrence of 'now' in it²⁶ - is highly significant. It points to the fact that internalism is a transcendentalist doctrine, since the fact that any particular dated moment is now is a transcendental fact. I just lodge these claims here without elaboration: I shall in due course (chapter 3) be devoting considerable attention to the transcendental status of the now, and I shall return there to the significance of the above claim. I merely record here that internalism depends upon that transcendental status, and so awaits a elucidation of that status for its legitimacy.

Still, without broaching that issue, we can at least say that the cost of adopting externalism is that consciousness is no longer diaphanous or transparent to the conscious subject: he is no longer an authority on the question whether or not he is conscious. But it is not clear that consciousness without transparency is an option

for us: it would certainly be a very hard doctrine to embrace. Perhaps we could still find consciousness in a world without incorrigibility, but it would not be the world we thought we lived in. And that fact in itself seems to be a problem for the externalist: how is it that we were ever tempted by internalism? It will have been noted that I have committed the manifest image to internalism. The supervenience thesis, for example, to which I have connected the manifest image, is not embraced by a radical externalist; but it does seem to me to be plausible that our image of ourselves incorporates the view that the subject has authority over the objects of his own states of seeming.

I should accordingly be inclined to say that the manifest and scientific images are in contradiction with one another over this issue. I would say that a world in which a subject has no authority over what seems to him to be the case would be a world without consciousness, as the manifest image conceives it. The defender of that image, whom we may as well call the 'Cartesian', wants to say: it is an a priori truth about consciousness that the subject is an authority on how things seem to him to be. The externalist denies that seemings have this status, and perhaps has then to deny too that there are such things as seemings, and so such a thing as consciousness; since perhaps the Cartesian is right that we cannot separate the idea of a subject's special authority from the concept of a seeming. The externalist will deny that we cannot make the separation: he must do so, since he certainly wants to say that subjects see, and plainly the concept of a seeing is not available independently of the concept of a seeming. But then how can things

seem to a subject to be thus and so, without his being an authority on how they seem to him to be? How could we abandon this connection without abandoning our image of ourselves as beings to whom things seem at all? It is, as has been noted, a further consequence of externalism that I cannot have a seeming without also having a seeing: I cannot have a mere seeming. For if I could, who on earth could have authority on its content if not I? This consequence too makes radical externalism an extremely counterintuitive doctrine.

If it is correct that 'to the extent that the manifest does not survive in the synoptic [the scientific] view, to that extent man himself would not survive' (Sellars [2], p.18), isn't that just to say that the scientific image cannot subvert the manifest? For how could we be surer of the correctness of the scientific image than of our own existence as men who enjoy experiences which are as the Cartesian says they are? How could I let the scientific image subvert my own belief in myself? And would its effect not be subversive, since surely I am a being who (as Sellars puts it) has a certain image of himself, and does this image not essentially incorporate a belief that I am authoritative on how things seem to me to be? How could I survive the subversion of that authority? Am I not partly constituted by it?

Chapter 2: Objectivity and Secondary Qualities.

2.1 The Causal Ground Argument.

In chapter 1 I denied that causality plays a role in hooking up the subject to the experienced world. That conclusion did not, as we saw, worry the radical externalist, who merely relocated the causal transaction away from the experienced world and mind to the scientific world and brain. But in the context of a certain attractive internalism - that the subject has authority on how things seem to him to be, on which things seem to him to be that way, and in general on whether or not he is conscious - the anti-causal thesis has important repercussions for our view of the fabric of the experienced world. One obvious corollary is that the causal domain can place no restrictions on what sorts of world of experience may be available to a subject. If the only relevant thing that we can say about the nature of the experienced objects of a world is that they are essentially such as to be capable of figuring as the content of representational experience, and if the subject is an authority on whether or not he has such experience, then nothing on which the subject has no authority - such as the causal provenance of his experience - can have any say in what gets experienced (what gets into a world) and what doesn't. I shall

suggest that one way of putting the claim is this: only secondary qualities are inherent in a world. The primary - if and where it is available at all - plays no role in constituting the objectivity of a world capable of being experienced by perceiving subjects. In chapter 1, I distinguished between representational experience simpliciter, which presents its subject with a world, but not necessarily with the world, and indefinitely corroborable representational experience, which presents its subject with the world. In this chapter I shall not be concerned with the distinction between veridicality and error in perception, but with that between objectivity and subjectivity, with what it takes to constitute representational experience (i.e. experience of an at least inchoately objective world). I shall accordingly be concerned with representational experience in general, whether or not indefinitely corroborable.

I am going to offer a further elaboration and defence of the world-view so far promulgated in the context of a consideration of the distinction between primary and secondary qualities. The opposing view of the distinction to the one which I shall take, canvassed for example by Evans in his article 'Things Without the Mind' ([1]), is that primary qualities have a special role to play in constituting a subject capable of experiencing an objective world. The initial idea (no doubt acceptable in itself) is that one cannot simply help oneself to a representational phenomenology: such a phenomenology has to be earned (i.e. constituted). A purported subject of merely secondary experience would not have the resources, Evans suggests, to enjoy experience of an objective world; it takes

experience of primary qualities to constitute the objectivity of the experienced world.

The traditional view of the primary/secondary distinction is that it corresponds to an objective/subjective distinction in a somewhat different sense from this: secondary qualities are thought to be mind- or experience-dependent in a way primary qualities are not. Even if Evans' argument for the primacy of primary qualities in constituting a subject of experience of an objective world fails, this view may be able to stand. On this view, the purported experience-independence of primary qualities is thought to guarantee their status as parameters which 'will always pull their weight in all explanatorily adequate theories of the world' (Wiggins [3], p.120). I shall argue that both the view that primary qualities constitute objectivity for a subject, and the view that primary qualities constitute objectivity inasmuch as they and secondary qualities differ in point of experience-dependence, and they are explanatorily more powerful, are mistaken. My conclusion will then be that objectivity - the world - is fully constituted by the secondary.

Let us start with a rough demarcation between primary and secondary that we can work from. Let us say that primary qualities are properties which figure in an a posteriori physical theory. This is vague, but it will do to be getting along with. They are such properties as mass, electric charge and energy. Also under this heading we must include geometrical properties of matter and constituents of matter (such as the shape of a table or a molecule), but not geometrical properties simpliciter, or as instantiated by

secondary qualities (such as the shape of a colour patch). This ambiguous attitude which the tradition has towards geometrical properties is a problem to which I shall recur. Secondary qualities are experienced properties which do not figure in a physical theory, or do not need so to figure (that is not what they are there for). They are such qualities as colours, tastes and sounds. As I am setting up the terms, primary qualities are proprietary to the scientific image, secondary to the manifest. It may seem surprising that I have demarcated primary and secondary qualities in this way. Traditionally, it has been customary to define secondary qualities as dispositional properties of objects in the following sense: for an object to have such a property is for it to be such as to give rise (in suitable circumstances and in a suitably conditioned subject) to an experience of that property. Primary qualities are then defined as properties of objects which are not dispositional in that sense.¹ I am not following this tradition because I want to raise the question whether primary qualities as traditionally defined are instantiated anywhere, and if so where. But in the meantime I need a terminology which enables me to discuss the question whether the properties which have been thought to be primary in my sense have a special role to play (1) in making objectivity available to subjects of experience and (2) in comprising the principal subject matter of explanatory theories of the world. It is convenient to call these properties 'primary', leaving open the question whether they fit the traditional definition. I shall tackle the second of these questions in the *third* section. In this section ^{and the next} I shall examine the first question in the

context of a close reading of Evans' argument. Evans just assumes that primary qualities as I have fixed them fit the traditional definition. I do not need to question his assumption at this stage (I shall do so in 2.3); the ensuing discussion will accept it pro tempore. What I am now trying to ascertain is whether in order to have experience of an objective world, a subject needs to have experience of spatial properties of enmattered objects at all, irrespective of how he thinks of these properties. An objective world is, in this context, a world not constituted by experience of it in this sense: it is a world which can be thought of as there in the absence of experiencing subjects.

Evans argues that one cannot form a conception of an objective world on the basis of experience alone, not even primary quality experience: one needs in addition to master an elementary theory of primitive mechanics, which will perforce be couched in primary terms. Consequently, a subject of experience who perceives only secondary qualities, such as sounds, is not in a position to develop a conception of an objective world. We may, following Evans, set up such a subject and call him 'Hero'. In his reply to Evans ([5]), Strawson imagines a subject who has only visual experience ('Seer'). Now, as far as Evans is concerned, the main difference between Hero and Seer is that Seer can, as Hero cannot, perceive primary qualities: so he is at least in a position to form a conception of an objective world. But that does not of course guarantee that he will avail himself of his opportunity.² To achieve this conception Seer has not merely to experience primary qualities, but also experience them as such, that is to say (as Evans assumes) as

essentially non-experiential properties whose identity conditions are fixed by their role in a theory of primitive mechanics. What Hero needs, but cannot have as long as he is confined to a world of audibilia, in order to be able to give objective content to his experience, is the notion of a categorical grounding causally responsible for his experiences. And categorical, Evans takes it, means primary. Suppose you look out of your window and see a street violinist ^{the sound of whose bow is} drowned by a passing band:³ the claim is that your ability to perceive the movements of the bow - and to see them as essentially non-sensory phenomena - grounds your ability to think of the violin's sound as there to be heard, although currently unheard because of the noise of the passing band. Strawson himself, in his discussion of the case, makes only the weaker claim that the persistence of the causal base provides evidence justifying a reasonable belief in the existence of the unperceived sound; but Evans thinks (p.105) that it constitutes our ability to form such a belief.

The substantial claim is not that you need the concept of the categorical for the objectivity of what you experience - that is trivial. (What is meant by the requirement of categoricity is that the subject should conceive of some of the objects of his experience as having an existence independent of his particular experience of them: and that is just the objectivity requirement.) Rather, it is that we should look for categoricity at the primary level; and that the contrasting phenomenon, mere dispositionality (i.e. dispositionality without categoricity), is to be met with at the secondary level.

But if we allow Hero to have experience of sounds, has not the trick been turned already? Hasn't he thereby achieved a conception of an objective world of sounds? Evans sees this objection coming in a related context: he has allowed Hero serial or travel-based spatial concepts, that is to say, concepts of temporally related experiences (e.g. an experience of F coming before an experience of G and after an experience of H). He then imagines the objection that seriality is enough for objectivity, and replies:

Because serial spatial concepts do not provide us with a way of thinking about simultaneously existing objects, they are not obviously concepts of relations between (independently existing) objects at all. (pp.113-4).

But this reply leaves matters in a very unsatisfactory state. If serial concepts are not concepts of objects, what on earth are they concepts of? Evans himself does not respect his own claim, since he writes of Hero's serial concept 'x is between y and z':

What it means to say that x is between y and z is simply that an experience of x will intervene between any experience of y which is followed by an experience of z, and conversely (p.110).

But if these experiences are not, strictly speaking, of anything, what are they? As things were set up, Hero was supposed to hear sounds: there was nothing in the official characterization to suggest that, just so far, his experience differs from our experience of sounds, except in point of spatiality. We hear sounds disposed in public 3D space, but Hero's experience of sounds, if it is spatial at all, is in the first instance experience of sound in a

1D quasi-space. Hero's experience of sound is to be equated with that of one of us listening to music through headphones: if the sound has spatiality at all it is certainly not the spatiality of our ordinary experience of sounds, which is experience of sounds as disposed in (at least) an egocentric space. But the space of sound heard through headphones, and of Hero's sounds, is not even an egocentric space: it is completely disjoint from the space of ordinary 3D objects. Nevertheless, experience of sounds heard through headphones is, precisely, experience of sounds - there is simply no existentially less committed way to characterise it. Any characterization of it which failed to imply the real existence of the sounds heard would quite simply falsify the phenomenology. Or so the objector is claiming. (I shall come to the question of what it would take to spatialize Hero's world in due course. For the moment we need assume no more than that he hears an unordered succession of buzzes.)

So far we have not considered the question whether Hero can be credited with some of the abilities commonly thought to be involved in objectivity, and which one might suppose needed to be incorporated into Hero's phenomenology if it is to earn representationality. These include abilities to

- (1) reidentify particulars, or at least distinguish in thought between the recurrence of the same particular and the recurrence of the same universal but a different particular,
- (2) conceive of perceived and unperceived objects existing simultaneously and in exactly the same sense (the simultaneity and topic-neutrality requirements),

- (3) distinguish successive perceptions of simultaneous existents from successive perceptions of successive existents, and
- (4) conceive that one's perception might be mistaken.

We may add to our list of purported marks of objectivity another, which is crucial to Evans' argument, and which I touched on earlier. This is (5) the Causal Ground Requirement. For experience to be constituted as experience of an objective world, it is necessary, according to Evans, that the experience be at least partly directed onto features of reality that are (a) not essentially experiential, (b) ground experience causally and (c) are conceived by the experiencer as possessing these first two characteristics. But if the objector is right to find objectivity built into Hero's experience as he has been set up, then that motivates him to reject any of these requirements should they not be incorporable into Hero's world. In advance of a consideration of these purported requirements, the objector claims that we must, on pain of falsifying Hero's phenomenology, characterise his experience in objectivist terms.

An important restriction on the scope of this objection must be entered here. The objection assumes, with Evans, that there is some sense in which Hero does have experience, and some sense in which he is able to conceptualize that experience; the point will then be that Hero's experience cannot fall short of a fully conceptual experience of an objective world. I am going to accept the objection. But then the dialectic cannot simply dissociate itself from the question of the status of the five purported marks of objectivity. For if any of these marks is constitutive of experience

of an objective world, and if further it cannot be fitted into Hero's world, then it follows that the terms of the thought-experiment are not coherent. I shall in fact argue that all the marks which could conceivably be constitutive can be fitted into Hero's world, so that if I am right about this the thought-experiment will remain intact. But Evans supposes that even if Hero does not have experience of an objective world, he does at least have experience of some sort, and it is this joint supposition which the objector attacks. Given that the objection will be upheld, the question of the coherence of the thought-experiment itself will have to await the detailed consideration of the purported marks of objectivity. What is argued for here is just the conditional: if Hero has experience at all (if the terms of the thought-experiment are coherent), then it is experience of an objective world.

So the question we have to ask is: has the objector got Hero's experience right? Evans thinks not. But it is clear from many places in the article that he conceives of Hero's experience of sound, and hence our experience of sound as heard through headphones, empiricistically (and it is hard to see how anyone who took issue with the objector could do otherwise). In one place (p.96), he writes that

it is not possible to distil the concept of hardness solely out of experiences produced by deformation of the skin which is brought into contact with a hard object.... [it is not possible] to have a purely kinaesthetic conception of what it is for one's legs to be crossed....and, although this is less obvious, it does not appear to be possible to regard the conception of a shape of a material thing as the same as whatever shape concept might be grounded in the colour mosaic thought to be given in immediate visual experience.

Evans is more tentative here about the visual case, but even so his use of the phrase 'immediate visual experience' indicates that he is operating with an empiricist notion of experience in this case as in the others: something which is available to the subject in advance of being imbued with objective content. Later (p.103), Evans argues that Hero's experience of sounds and the sounds themselves are separable in this sense: Hero can have the former without our being required to admit the existence of the latter. We can now fit this claim into a larger empiricist framework underwriting Evans' thought here. As a final confirmation of this, in a third place (p.114) Evans allows the possibility that Hero's experiences should be subject to a 'simple phenomenalist reduction'. For this to be possible, it must be the case that the content of Hero's experience is conceived to be exhausted by purely intrinsic (i.e. not inherently world-directed) features of that experience.⁴ The suggestion is that we do not need to find a place for Hero's sounds, because we do not need to objectivise them. Statements purporting to ascribe to Hero experience of independently existing sounds are reducible to statements containing no such purport.

Let us recall that Evans allowed Hero serial but not simultaneous spatial concepts. The objector questioned the point of this move on the grounds that confining Hero to seriality in his experience isn't going to damage the prospects for objectivity of that experience, if that is the way its phenomenology pushes us. Even if Hero's experience is inescapably travel-based, if it is still unavoidably experience of sounds, then it has objectivity built into it. Now the objector's point can be illuminatingly put in

the following way, and realising this will enable us to see our way to upholding his position against Evans'. The concepts in play here ought to be Fregean concepts; that is to say, concepts essentially of objects. (Objects are just whatever Fregean concepts are concepts of.) So if Evans wants to stop Hero objectivising from the very beginning, he ought not to allow him serial spatial concepts, but rather just serial spatial experience. And the experience in question here must, as we have seen, be read empiricistically. We can now put the gloss in this way: Hero's experience must, for Evans, be non-conceptual. His experience has a travel-based element, but is essentially unconceptualised.

Now that we have got clear on this, I think it is possible to see why Evans locates categoricity in the primary, and hence demands primary quality experience for objectivity. Experience of secondary qualities, which Hero has, is not enough, because experience of secondary qualities is not, just so far, conceptualized, and hence not yet experience of a genuinely independent reality. The conceptualization, and hence objectivity, come when a subject also experiences primary qualities, and experiences them as such - as not essentially experiential qualities - and conceives of them as grounding experience causally.

We come back, then, to the Causal Ground Requirement. But Evans' inclusion of the Requirement is, as we now clearly see, consequential on his construal of experience as, in the first instance, non-conceptual. So the Requirement does not hang separately from an entirely controversial empiricistic conception of experience. We should represent Evans' argument as proceeding via

the contrapositive of the following principle:

(P) If a subject (e.g. Hero) has a conception of an objective world, then he conceptualizes his experience;

assuming the supplementary principle:

(Q) To conceptualize one's experience is to conceive of it as grounded in phenomena that are not themselves essentially experiential.

Hero, not being in touch with primary qualities, cannot regard his experience as thereby grounded, and so cannot conceive of an objective reality. The insertion of the notion of conceptualization into the argument, and the gloss on 'experience' as 'non-conceptual', are necessary to make sense of it, and provide some bulwark against the objector we mentioned earlier.

Some bulwark, but not, in the end, an effective one. For it is open to us just to refuse to accept the empiricist notion of experience. We can insist that experience is, essentially, conceptual. I defended that insistence earlier (1.1), and this is one place where I lean heavily on that defence. I proceed on the assumption that we can count on experience consisting, all the way through, of conceptual structures.⁵

2.2 The Marks of Objectivity.

Where does this insistence lead? Well, it pretty clearly leads to the collapse of the Causal Ground Argument. The point of the Requirement was to give independent existence to a realm of objects of experiential states. Experience itself cannot provide the necessary certification if it is construed as a purely intrinsic affair, whose content is exhausted by purely intrinsic features. Such an experience would not have the resources to equip a subject with a distinction between states of the experiencing subject and states of the world. And that is how Evans, in taking experience to be, in the first instance, non-conceptual, construes our and Hero's experience of secondary qualities. Hence the Causal Ground Requirement, and hence, if we reject the notion of experience as being, in the first instance, non-conceptual, our entitlement to reject that requirement. So the Causal Ground Requirement, which was our fifth purported mark of objectivity, falls.

If there is no point at which experience is non-conceptual, and waiting to be jumped up to the conceptual level, then there is no justification for a requirement which would make experience of secondary qualities parasitic on experience of primary qualities: experience of secondary qualities comes ready-made; it does not have to be constituted out of experience of primary qualities, plus a bit of elementary physical theory. Neither empirically constituted nor transcendently constituted: in neither the empirical nor the transcendental mind is there any justification to find the raw material - namely non-conceptual experience - out of which the

constitution is supposedly effected. (It might be thought that an analogue of a primitive physics had to be present at the secondary level - with, say, the mutual exclusion of colours providing an analogue of the mutual impenetrability of material bodies. I shall not consider this view here, but I shall return later to the role of geometrical properties in primary and secondary quality experience.)

As so far described, Hero hears sounds in a way resembling the experience of listening to randomly produced noises through headphones. If now we allow, as I think we have overwhelming independent reason to do, that his experience of sounds is, like ours, conceptual, then we are forced to agree with the objector that his experience of sounds really is experience of sounds, that is of audible phenomena enjoying an existence potentially independent of the perceptual episode. This just unpacks the nature of the phenomenology which we have built into Hero. We have given him a representational phenomenology, and assuming the coherence of the thought-experiment that phenomenology is not available separately from the potentially independent existence of the experienced objects. (Of course in setting up a subject of bare audibilia at all, we had to make the phenomenology representational - there is no such thing as the non-representational experience of sounds.)

There is an important principle at work in this line of thought; we can help bring it out, as well as discharge our obligation to support the coherence of the thought-experiment, by a consideration of the first four purported marks of objectivity. This I shall now attempt.

What of the first purported mark, reidentifiability of

particulars? Clearly there is no problem about Hero's ability to reidentify sound-universals, but it might be thought that Hero cannot have the concept of a sound-particular, because he can't distinguish in thought between the recurrence of the same universal and same particular, and that of the same universal but a different particular. Certainly we should reject Strawson's attempt to achieve this by equipping Hero with a master-sound. Strawson supposes that the notion of the same sound-universal heard against the same pitch of the master-sound might add up to a notion of the same sound-particular, but the introduction of the master-sound does not so force us to find persistence of particulars in Hero's experience: it can still all be captured in terms which do not presuppose that the recurrence of the same universal is ever the recurrence of a particular.

What is a sound-particular for us? It might be thought that to generate this notion one had to appeal to the primary base: perhaps a sound-particular is the sound universal produced by a particular primary base (this is what Evans supposes). But I do not think this is so. Imagine a note of constant pitch, dynamic and timbre being played by a succession of different violins, in such a way that the transition from one violin to the next is quite imperceptible. We should surely want to identify the sound-particular here with the whole performance. A change in the primary base as we go from one violin to the next does not affect the continued existence of the same sound-particular. Reflection on this sort of case suggests that sound-particulars should be individuated by their temporal parameters (as well as their universal features), and not by primary

base. If that is right, it is clear that we can equip Hero with a conception of sound-particulars, since he is obviously able to discriminate discrete stretches of sound-universals.

But can he reidentify a sound-particular? What we want to be able to do is equip Hero with the capacity to conceive that this sound heard after an interval is the same particular as that sound heard before it. We do not necessarily need to equip Hero with a practical ability to reidentify sounds. It seems that we can incorporate the conceptual ability via the possibility of varying his receptivity. If we give Hero the possibility of falling asleep and waking up, or of losing and regaining concentration, and if we build some self-awareness into his phenomenology - so that he feels himself falling asleep, or regaining concentration - then he can achieve a grasp of the notion of sameness of sound-particular. And he can use the notion of a lapse in receptivity to explain to himself why, on some particular occasion, a sound which was there to be heard was nevertheless not heard by him.

Hero would not indeed be in the same position as we are in respect of sounds, for we can verify that this is the same sound again by verifying the persistence of the same primary base (which is a sufficient, though as we saw not a necessary, condition of the persistence of a sound-particular); but he would be in the position we are in in respect of, say, headaches. I say that I wake up with the same headache I went to sleep with, but I have absolutely no way of verifying this. Perhaps the time will come when it is possible to verify such a speculation (although it will remain in principle impossible to falsify it), but it would be fantastic to suppose that

for centuries people have, in conceiving their headaches to persist unfelt through their sleep, been supposing that the primary base of the pain has persisted. Rather, what is conceived is that the pain persists, and one's only grip on this is that one fades out on the pain, and later fades back in on something qualitatively indistinguishable from the original pain. The thought involved in supposing that my headache persists through my sleep is just this: had I at any point woken up, I would have felt a headache (with the relevant quality). The fact that it is I who fade on the pain, and not vice versa, is essential to the ability of a pre-scientific mentality to conceive of the pain as persisting unperceived. But not only essential, also adequate. And if adequate for our headaches, then adequate for Hero's sounds.

It is important to insist on respecting the phenomenology of receptivity here: Strawson and Evans both claim that we cannot provide Hero with a way of distinguishing between his fading on the world and the world fading on him, purely in terms of his varying receptivity. There is a difference between their treatments here, which we should note. Strawson objects to the auditory world as such on the grounds that in it no basis can be given to the above distinction. Evans allows that sense can be made of the distinction in a spatialized auditory world (i.e, one in which sounds occur with sufficient order and regularity for the subject to construct a map of his world: see below for more on this). This is because the circle connecting world and receptivity is not immediate (Hero can always rely on other parts of his world in judging that he was, on some occasion, unreceptive). Evans just objects to the idea that

Hero might rely entirely on his receptivity to distinguish the world being thus and so and his being unreceptive to it, from the world not being that way. Hero must be able to get independent empirical grip on his receptivity: the spatialized world provides that independence, and the suggestion is that receptivity on its own cannot.

Now we should accept the point that Hero's grip on his receptivity must be independent of his conception of what, on any particular occasion, is actually perceived. Providing Hero with a map of a spatialized auditory world is certainly one way of supplying the independence, but it is not the only way. An alternative is to enrich the phenomenology of receptivity so that Hero can simply perceive that he is falling into, or coming round from, unreceptiveness. He does not land himself in the incoherence of relying on what he perceives to determine whether or not he has been receptive, and then circularly trying to depend on his receptivity to tell him what was there to be perceived. He can get his empirical grip on his own receptivity not from what he experiences, but from the phenomenology of his experience itself. Hence we can equip Hero with a concept of sameness of sound particular, even if he can't in practice reidentify. And we have again seen the importance of respecting phenomenology in doing this.

The enriching manoeuvre provides us with material to construct what I have already argued is a constitutive part of a conception of an objective world, namely a simple theory specifying enabling conditions of perception. Recall that a subject of experience has to be able to make sense of the idea that the very same kind of state

of affairs is capable of existing perceived or unperceived. If, and only if, he has such a theory, the subject can make sense of the idea of a state of affairs, which he reidentifies after an interval, persisting unperceived through that interval. He does so just by entertaining the thought that those conditions were not met. (That thought cannot be tantamount to the thought that he did not perceive what was there to be perceived, because then the circle between world and receptivity would be too narrow: perhaps there was nothing there to be perceived.)

Spatializing the auditory world is one way out of the difficulty: that way widens the circle. But enriching the phenomenology is another equally good way of achieving our aim, and it works by breaking out of the circle. On this manoeuvre, we appeal to one of the enabling conditions - receptivity - to make sense of unperceived existence, and we are entitled to do so because Hero's grip on his own receptivity is completely independent of the world. He finds out how his receptivity is varying by asking himself how he feels. The first purported mark of objectivity can accordingly be fitted into Hero's experience, and we have seen that it can be earned without spatializing his auditory world. We have also seen that this mark is constitutive of objectivity, because it does not hang separately from the simple theory of perception which every subject of experience must possess.

The second purported mark of objectivity incorporated simultaneity and topic-neutrality. There is no difficulty in awarding the latter conception here to Hero. We do so in respecting the representational nature of his phenomenology, for that

phenomenology guarantees that experience of a reality, even if it is serial, actually forces the experiencer to think of the experienced objects as existentially on a par.

The former requirement is that the subject be able to conceive of perceived and unperceived objects existing simultaneously. Now our ability to do this depends on our ability to locate things at places. As Strawson says,⁶ we both need and have a dimension other than the temporal in which to house unperceived particulars. And it would appear that without providing Hero with an analogue of space we cannot meet this challenge. If Hero simply perceives a succession of unordered sounds, then Evans is correct to say (p.86) that if, at any time, sound of kind F is perceived, then that is all the sound of kind F the world affords. In fact we can say something stronger than this, in the light of our previous discussion on the phenomenology of receptivity. If, at any time, Hero hears some sound of kind F, then that is all the sound that he - as so far described - can conceive of as existing at that time, given his awareness that he is, at the time in question, receptive. So in the unspatialized auditory world, the simultaneity requirement could not be met.

Let us then spatialize it, and see what happens. We can give Hero a symphony, say, which he can mentally engage in and listen to. We can also equip him with fast-forward, rewind and play facilities under his mental control. There will be a holistic interdependence, for the instantaneous chords of Hero's symphony as for our material particulars, between things and places. I think Hero as he is now constituted is actually forced to think of the parts of his symphony

as existing simultaneously, in an entirely parallel way to the way we are forced to think of material bodies as existing simultaneously. We, of course, can conceive of a sound, which sometimes exists at a place p , as not always existing there, because we can conceive that there is not always a suitable primary base at p , and because we do not individuate places by sounds; but Hero differs from us in both these respects. Hero as he is now constituted is forced to think of the instantaneous chords of his symphony as existing simultaneously, because, if we imagine him listening to the first movement of his symphony, he is not able, as we are, to conceive of the primary base of the third movement as not yet activated (the relevant stretch of tape not yet over the tape-head). That is because Hero has no primary concepts at all: Evans thought that this ruined his prospects for inhabiting an objective world of sounds, but we have rejected that line. In fact what it means is that the sounds of Hero's symphony, having been found places to reside at, cannot but be conceived to exist permanently at those places, and hence cannot fail, all of them, to exist simultaneously. So how does Hero think of his symphony? He must think of it laid out quasi-spatially, so that all its instantaneous chords are being played simultaneously and repeatedly. When he mentally winds forward, he passes quickly down the line of orchestras each playing their assigned instantaneous chord; when he uses the play facility, he passes the instantaneous orchestras at the rate of one orchestra per instant.

Hero has, in fact, a cognitive map of his symphony. There is an interesting feature of his auditory world thus spatialized, which

I do not think upsets its coherence, but which should not go unremarked. Suppose we equip Hero with a collection of different symphonies, any one of which he can select for listening by, say, pronouncing its name to himself. This addition to Hero's world effects a partitioning on it into a set of mutually disjoint spaces, one per symphony. For although universals are instantiated across the board, particulars are symphony-bound. This is because the identity conditions of particulars are given in terms of their quasi-spatial relations to other particulars within their respective symphonies. Symphonies are for Hero what dream worlds are for us. There are only universal links between them.

We saw that the first mark is constitutive of experience of an objective world, and can be fitted into Hero's world without spatializing it. The second mark calls, as we have seen, for spatialization, but there is no difficulty about supplying this. Is it constitutive (as Kant thought)? I am inclined to say that topic-neutrality is constitutive, but that simultaneity is not. I do not see any objection to our imagining Hero inhabiting an unspatialised world of sounds, under the principle that there is only one audial place: as long as Hero is receptive, the sound (if any) which he hears is all the sound there is.

As for the third purported mark of objectivity, we can again fit it into Hero's world by a simple appeal to his phenomenology. The question is: how can he distinguish successive perceptions of simultaneous existents from successive perceptions of successive existents? This is a question raised within an empiricist framework, and the reply is to eschew that framework. We should insist that

there is no level at which one can have anything amounting to experience which does not itself declare whether it is experience of simultaneity or of succession. The ability to make this distinction does not have to be constituted out of more primitive experiential material. The sort of attitude to representational experience which would find this problematic (essentially Kantian: see the Second Analogy) is epitomized by the following remark of Leslie Stevenson:

When one does decide that there has occurred an objective change, this is not implied merely by the temporal order of one's judgments; so there must be some other criterion by which we can know the objective ordering of events (p.52-3).

The suggestion here is that the objective order of events could at best only be implied by the ordering of one's experiences, and that in the absence of such an implication one stands in need of some criterion to take one from one's experience to the world. But the picture of the subject trapped behind his experiences is just what we leave behind when we abandon empiricism. Representational experience (if corroborated) takes one all the way to the world. The world just is the content of corroborated representational experience.

One reason why people have been led to deny this is that they have been impressed by such phenomena as the following: one can hear a bat striking a ball some time after one sees it; one hears the thunder after one sees the lightning. It is tempting, when confronting such phenomena, to suppose that there is a dichotomy between the order of events in the world and the order of experiences, that the subject needs to make personal sense of this dichotomy, and that he needs to integrate the time-orders, if he is

to achieve a conception of an objective order of events on the basis of a merely subjective order. The suggestion then will be that the subject needs a rudimentary theory - obviously a causal theory - which when combined with a description of the order of events, will enable him to explain the order of his experiences. So finally, equipped with the theory and a description of his experience, the subject will be enabled to extrapolate back to the objective order of events, on the assumption that the actual order of events is that order which, when combined with the theory, makes best sense of the subjective experience.

The view I am criticising inserts the causal relation between the world and subject, not between items in the world. Kant, of course, from whom this line draws its inspiration, argues for the constitutive status of the latter sort of causal relation. But we need a more general approach here, because we are looking for an explanation of the subjective order in terms of the objective order, and such an explanation is obviously going to have to go beyond a merely objective causal order. Now given my expulsion of causality from the theory of perceptual knowledge, I am also committed to rejecting this suggestion. What is wrong with it is precisely the dichotomy from which it takes its cue. For given the representationality of the subject's experience, and the non-contingent, specifically non-causal, connection between the world and his experience, it follows on my account that there can be no dichotomy between subjective and objective orders: the objective order precisely is the order of experienced events as they are experienced, because the world just is the content of corroborated

representational experience. And since the corroboration takes place over time, it can only affect questions concerning the existence of experienced items, not their order. It would be impossible to corroborate order: there would need to be a meta-time-order in which events in the experienced time-order could be ordered. The point is that if we extrude causality from our theory of knowledge - everything turns on this - then we must reject the idea that a causal theory is required by the subject to constitute his experience as genuinely representational.

An important corollary of my account, as I have already stressed, is that the subject does not, in our example, see or hear the physical event which occurs when the bat makes contact with the ball: he sees the sight of the contact, and after a delay (if there is one) he hears the sound of the contact, and that is all. The world of experience contains here a visual event followed by an audial event. These events are caused by, and not identical with, the original event of the bat striking the ball. (Why the audial event follows the visual event is something we need a bit of physical theory to explain.) We do indeed talk of seeing the bat strike the ball, but we also talk of seeing the sight of the bat striking the ball. I claim that the first locution is an ellipsis for the second. I am committed to saying this by my theory of knowledge; to say that I saw or heard the physical event of the bat striking the ball would be to say that I could infer, from my experience, that such a physical event had occurred. But of course no such inference would be legitimate. I can infer the existence of intentional objects of my experience (sights and sounds), but not

real objects, let alone physical correlates of real objects. The battle between the world-view I am offering and the one I am opposing turns essentially, as I have said, on the issue of causality. My view here involves an internalist and anti-causal conception of what the world we experience is; if there is another world, causally underpinning this one, that is no business of subjects of experience. It is a consequence of this approach that there is no problem about integrating subjective and objective time-orders.

Finally, I consider the fourth purported mark of objectivity: the ability to accommodate perceptual error. There is little to say here: we have already incorporated this ability into Hero's world in equipping him with the first ability (to reidentify). We saw that this ability was consequential on the simple theory of perception which must be ascribed to any subject of experience. There was no difficulty in awarding Hero that theory: we gave him a conception of those enabling conditions which are subject-dependent by varying his receptivity and building self-consciousness into its phenomenology; and clearly there is no difficulty in giving him a conception of those enabling conditions that are world-dependent: he can simply hear distortion or interference. These points carry easily over to the fourth purported mark: we just need to observe that the simple theory is, as we saw in chapter 1, precisely what provides for the possibility of perceptual error.

This concludes my discussion of the five purported marks of objectivity. We have accommodated the first four marks in Hero's world, but we have rejected the attempt to ground experience of

secondary qualities in experience of primary qualities. If the first four purported marks are genuine marks, as the fifth is not, we are then able to say that the Heroic thought-experiment is coherent, and hence that it is possible to imagine objective worlds composed of such phenomena as sounds.

Now, given the failure of the grounding connection, we have it that, contra Evans, it is conceivable that there should be a world whose ontology consists entirely of secondary qualities; secondary qualities are in fact - in just the sense Evans wanted to deny - categorical. This goes for our sounds as much as for Hero's. The connections of mutual implication between sounds and primary bases is an a posteriori matter, and doesn't affect the phenomenology of sound experience. That phenomenology forces us to ascribe categoricity to sounds.⁷ Now Evans accepts that mere dispositionality (i.e. dispositionality without categoricity) doesn't enter into the meaning of terms such as 'red'. But he nevertheless wants a sense in which ascriptions of secondary qualities are 'true in virtue of' the presence of certain primary qualities. And this sense is stronger than a contingent supervenience. It is an a priori relation between the two kinds of quality: Evans aims to offer a transcendental argument for the thesis that secondary qualities are objective only insofar as they are conceived as being grounded by primary qualities. And it is that argument we have seen to fail. We are left with an ordinary (empirical) sense in which red supervenes on the microscopic textures of objects, or the sound of a violin on the movement of the bow across the string. The difference between our position and

Evans' can be brought out clearly if we return to the street violinist. For Evans, the persistence of the primary base grounds the intelligibility of the persistence of the unheard sounds; for us, it merely grounds the truth of a belief in the persistence of that sound (via supervenience).

The difference between my position - that red is a categorical property - and Evans' - that it is a merely dispositional property - is this. Evans thinks that you can mean either of two things by 'red': you can mean the primary base, or you can mean your own experience, empiricistically construed (p.95). You cannot mean - what I claim you both can and must mean - an experiential but withal categorical property of objects. Evans attacks an opponent who tries to say

referring neither to the experience nor to any primary property of the thing, 'this, just as it is, can exist in the absence of any observer.' (p.98)

The difficulty with this is that one's experience of colour is already imbued with objective content, so that it is perfectly in order to concentrate on one's experience of red and thereby come to know what it is for an object to be red in the absence of any observer; for concentrating on one's experience of red will, in the most favourable cases, precisely involve looking at a red object.

This response must be distinguished from Mackie's position which Evans principally has in his sights in the above passage. According to Mackie,⁸ common sense attributes to objects colour properties which are not essentially experiential in mistaken response to perception of experiential such properties. This view

would make common sense guilty of the straightforward incoherence of wanting redness to be both essentially experiential and not essentially experiential. (It would want objects to be red in some stronger sense than that they look red.) But my response to Evans above is not implicated in this incoherence: I can accept that red is essentially experiential, while still denying that this makes red a feature of experience empiricistically construed.

If we take this line we can have both the essential experientiality of red and its categoricity (i.e. not mere dispositionality). Red is essentially seeable, but is an abiding property of objects in the sense that it can exist, in the absence of any observer, 'exactly as we perceive it'. Evans, as we saw, splits experience of colour into into an empiricistic subjective component, and the primary base. These correspond to the two limbs of the dispositionalist analysis.⁹ Now it is worth pointing out that although this analysis fails for secondary qualities like red and Hero's sounds, a version of it works perfectly well for objects of non-representational mental states, like pains. That is, either 'painful' refers to the primary base (mere dispositionality thus entering into its meaning), or it refers to the (non-representational) experience. There is no room for painfulness to be a secondary, and categorical, property of objects. (Of course the mere dispositionality in play here is not quite the same as that urged by Evans for secondary qualities: there need be no suggestion that the experiences which painful objects are disposed to give rise to are, in the first instance, non-conceptual.)

What is crucial in differentiating redness from pain here is

once again the phenomenological difference between the experiences in question. Colour experience, being representational, aims to be experience of what is there anyway: the experiencer does not intrude into the content of the experience. On the other hand, when I feel a pain, I feel myself in pain. Non-representational experience is not experience of what is there anyway, independently of the existence of the perceiver. This categorial divide between the representational and the non-representational cannot be bridged: in particular it cannot be bridged by Wittgenstein's so-called fantasy of pain-patches. Actually this is not a fantasy - electric fences constitute just such patches. But the existence of electric fences provides us with absolutely no temptation to assimilate pain to redness. The pain is in me, and not the wire. Evans supposes that his opponent should be prepared, if he wants categoricity for redness, to extend it to the painfulness of such pain-patches; but this supposition rests on a mistaken assimilation of the representational to the non-representational. If we resist that assimilation (see further 5.4), we can preserve the categoricity of the former while acknowledging the mere dispositionality of the latter.¹⁰

Evans would in fact need mere dispositionality at least to be able to enter into the meaning of 'red' in order to return an affirmative answer to the question: is a rose red in the dark? (Zettel, 250). We might sum up what is wrong with Evans' position by saying that he tries to answer empiricism by supplementing it, rather than abandoning it. He accepts the empiricists' account of experience as in the first instance unconceptualised, and then gives

you something else you can mean by 'red' other than your immediate visual experience - you can mean the primary base - and thus tries to secure the redness of the rose in the dark. Against this I contend that it is not possible to mean by 'red' either your immediate visual experience, or the primary base. Supporting this latter contention we have the possibility of variable realisation of the secondary at the primary level.¹¹ Supporting the former contention we have the conceptual and representational nature of colour experience. This phenomenological point is well brought out by Wittgenstein at PI 275:

Look at the blue of the sky and say to yourself 'How blue the sky is!'....If you point at anything as you say the words, you point at the sky.

Given that we are rejecting the merely dispositionalist account of the secondary - because we are rejecting its implicit empiricism - how can we secure the redness of a rose in the dark, and in what sense? To be sure one cannot imagine (picture) a red rose in the dark; anyone who claims he can is mentally turning on the light. But we can conceive of the redness persisting in the dark, and this is not to be glossed as conceiving that the primary base persists. Obviously you need light to see a colour, but to say that much is not to begin to concede Evans' dispositionalist picture, because on that picture what gives rise to the experience is not the colour itself, but the primary base. But saying you need light to see a colour is quite compatible with claiming that that very colour, which you need light to see, is there in the absence of light. (In any case I want to reject the causal overtones of the

locution 'gives rise to'.)

In fact we want to distinguish two claims one might make here, a true claim which my account is committed to, and a false one which it isn't. The true claim is that a red object is, in the dark, red, and red in a categorical sense. The false claim is that a red object is, in the dark, red-in-the-dark. This is false because nothing could have the colour property of being red-in-the-dark. That is because colours are essentially seeable, and redness-in-the-dark is clearly not seeable. This way of looking at the matter manages to capture our intuition that there is a sense in which a rose isn't red in the dark (namely: it isn't red-in-the-dark), but, equally, that there is a perfectly good sense, which isn't Evans' sense, in which a rose is red in the dark (namely: it is, in the dark, red).

Secondary qualities are, then, categorical. What grounds this categoricity? I think we should answer: strictly speaking, nothing. But the categoricity is, if not grounded, at least underwritten by the inherently conceptual and representational nature of the phenomenology of secondary quality experience. The general principle which has emerged from our discussion of Hero's case is this: phenomenology is a metaphysically fundamental feature of experience. All experience is conceptual, and an experience which is, in addition, representational, has irremovably built into it the inchoate objectivity of the experienced reality. If that experience were to be corroborated, the inchoate objectivity of the experienced reality would be upgraded so as to enjoy the status of the real. Of course a representational experience may fail to be corroborated,

but while such failure does confine the subject to the non-veridical, it does not consign him to the realm of the subjective, in any sense favourable to empiricism: it does not trap the subject behind his experience - what he lacks is more of the same, not something different in kind (or conceptualization of what he has). The objects of the uncorroborated representational experience have failed to achieve objective status, but not because the experience is ungrounded causally. Its character always was that it represented a world - potentially the world - and so it was potentially objective. To the extent that experience of the objective has to be earned or constituted, we saw that those purported marks of objectivity whose constitutive status could with plausibility be defended were incorporable into purely secondary quality experience. But we should remember that some of the argumentation here relied on the anti-causal thesis of chapter 1, and a certain plausible, but by no means uncontroversial, internalism. Hero has been our model of a subject whose experiences are such that the internalist's account of it is correct.

2.3 The Status of the Primary.

As I mentioned at the beginning of this chapter, there are two ways of viewing an objective/subjective distinction. Given that subjectivity is intimately connected to the having of a point of view, and objectivity to relative independence therefrom, the two ways could be said to differ in respect of their construals of the phrase 'point of view'. On one construal, a point of view is something enjoyed by an individual. The phrase is not intended literally: this sort of point of view is not to be identified with a particular point in space and an orientation: rather it is something the individual subject carries around with him; it is his individual subjectivity. On this reading, 'objective' will mean 'existing independently of the individual subject's point of view, not constituted by it'. On the other construal, the subjective is what attaches to experience as such, human or other (to the having of some point of view or other). The objective will then be the characteristic of phenomena whose existence is independent of experience in general. Independence here amounts to the property of being not essentially capable of being experienced.

Those who have argued that we need the primary to objectivize experience have had both of these construals of objectivity in mind. The contention is that not only for experience in general, but that even at the level of the individual subject, experience of primary qualities is needed to constitute any experience as experience of an objective world. The objectivizing works at the individual as well as the general level, because the

experience which the individual purportedly enjoys anterior to its objectivization is non-conceptual, and also necessarily private (in the technical sense introduced by Wittgenstein). So the isolated individual carries around with him non-conceptual experience of the secondary, and when this is suitably conjoined with experience of the primary, he can have experience of an objective world. This account embraces a subjective/objective distinction in the first of the above senses and so a fortiori in the second sense too. I have suggested that the underpinning of this contention is an unattractive empiricism, and that if we eschew it we can tolerate the possibility of constituting a subject whose experience is purely secondary, but is for all that genuine experience of an objective world. If we have the first of these construals in mind, the secondary turns out to be objective.

In this last section I turn to the second construal of the objective/subjective distinction, and argue that here we cannot straightforwardly correlate the objective/subjective distinction with a primary/secondary distinction as many have hoped. The question is: does the primary achieve the requisite degree of experience-independence for it to count as genuinely objective? If it does, then, assuming the equation between independence and explanatory adequacy mentioned at the beginning, primary qualities will be ideally suited to figure in what Bernard Williams has called an 'absolute conception'¹² or Thomas Nagel a 'view from nowhere'. I shall argue that they are not in fact unequivocally independent in the required way; but I shall also reject the implication from explanatory adequacy to independence (the converse implication is

evidently false); I shall substitute an implication from explanatory adequacy to dependence, so that I shall in effect be rejecting the notion that there could be an absolute conception.

It is important to see that our attack on Evans left undamaged a certain correlation between the secondary and the subjective in the second of the above senses. I conceded that red is essentially experiential, not in the sense of having actually to be experienced to exist (as the object of a non-representational state must be, at least for part of its duration), but in the sense of having to be capable of being experienced. We may call the version of the dispositionalist thesis favoured by Evans (the empiricistic version) the 'Strong Dispositionalist' account; the version I favour we may call the 'Weak Dispositionalist' account. On the Strong account, a primary base is disposed causally to give rise to an experience which is, just so far, non-conceptual. That account we have rejected. On the Weak account, a secondary quality (such as redness in an object) is itself disposed to give rise to an experience of it. The connection here is 'logical', not causal.

In general, the subjective in the second sense is characterised by a certain relation of dependence on subjects, and the objective is the contrasting notion to this. This dependence has been variously expressed: Hume talks of the mind spreading itself on external objects; we are familiar with the term 'projectivism' for the mental act involved here.¹³ But these locutions are not to the point. Projection requires a screen, and what is that to be? The idiom essentially allows for the possibility of alternative projections; how are we to accommodate these? With regard to the

former question, we simply do not have a coherent idea of what objects would be like anterior to the projection onto them of the mind-dependent phenomena: causal relations, moral and aesthetic attributes, constitutive connections, secondary qualities. How could there be objects lacking causal dispositions (see further 3.2), persons without value, objects in general without essences, or objects of experience without colour? We have no idea how objects could fail to be characterized in these ways. With regard to the latter question, we have no idea how we could effect the alternative projections onto them. The projectivist idiom carries with it an intolerable stowaway in the form of Dinge an sich.

Hence it is that projectivism (generally, idealism) must, as Kant saw, go transcendental if it is to be saved. But in that case the term 'projectivism' loses its content. There is no screen, and so a fortiori no projection." So we need some other way to register the mind-dependence which in some sense undeniably attaches to the items in the above list. The only way to do this, as far as I can see, is to talk about constitutive connections. What makes subjective features subjective is that they are essentially experiential (essentially capable of being experienced as such). They are not mind-dependent in the sense of having actually to be experienced, as Berkeley thought,¹⁴ but merely capable of being experienced (under suitable conditions). That is a weak enough connection, but it seems to be all we can legitimately salvage from the wreck of the Strong Dispositionalist account.

Now primary qualities, as I have fixed them, fall into two camps. There are posited phenomena, such as electric charge; and

there are the spatial properties of gross enmattered objects, such as the squareness of a table. Is either of these phenomena objective? Take first the example of electric charge. We must grant that charge is not essentially perceivable - it doesn't have to show up as such in a phenomenology (though it is going too far to say that it is essentially not perceivable); but it is, surely, essentially detectable, and I think this raises an insurmountable problem for its chances of figuring in an absolute conception. (I shall return to this shortly.) We do not have to espouse instrumentalism about fundamental physical parameters to see that the notion of such a parameter whose operation at the fundamental level did not show up on the observable surface of the world is not coherent. The positing of such phenomena is precisely geared to the project of explaining the behaviour of gross matter; so the detectability of such basic phenomena is really a logical truth.

Still, it will be objected, the presence of such phenomena does not have to be capable of showing up in a phenomenology. This we have agreed. So we agree that electric charge is objective in the second sense: it is not essentially experiential. But the requirement of detectability can amount to nothing less than detectability by subjects who are nevertheless in genuine possession of a phenomenology. Why is this? The reason is that a gadget or creature without a phenomenology could not be said to observe its instruments and react to them as instruments. Such a thing would be a mere extension of its instruments: on what basis could we distinguish it from them? The causal nexus would develop seamlessly from the world through the 'instrument'; then through the

creature/gadget and out beyond. Of course we might, for our scientific purposes, treat a part of this nexus as an instrument; but we would not be able to draw boundaries around another part of the nexus, call it 'the observer', and regard the first part of the nexus as functioning as an instrument for it. There would be no warrant for such a move. If this argument is sound, it would appear that such primary qualities as electric charge, although not directly phenomenological, in the sense of having to show up as such in a phenomenology, are indirectly so, in the sense of having to produce effects which so show up. So while such phenomena are objective, they are intimately connected with subjectivity. Their objectivity is not independently available. And given the internalism with respect to representational experience which I endorsed in the last chapter, it follows that the status of such primary qualities as electric charge is crucially beholden to a phenomenology which the scientific image is not competent to reconstruct. The scientific image is, at this point, in the service of a manifest image which transcends it.

Posited phenomena like electric charge are one sort of primary quality; what about spatial or geometrical properties as instantiated by enmattered objects or their constituents? What should we say about these? As regards geometrical properties of the unobservable constituents of matter, the account we are offering can simply carry straight over. But in the case of the geometrical properties of manifest objects, a quite different conclusion is called for. For the squareness of a table, say, must surely be counted in with its redness as a property which is subjective: that

is, it must be capable of showing up as such in a phenomenology. Of course, we must register a distinction here between the subjectivity of redness and the subjectivity of squareness. Squareness, although essentially experiential, has an objective component which redness lacks. A red object is such as to look red; a square object is such as to look as if it is square (i.e. square by the geometer's lights). Looking square has an objectivist commitment which looking red does not. This is reflected in the fact that disputes over what colour a thing is are merely verbal, whereas disputes over what shape a thing is can be settled by having recourse to measurement. Even the undoubted fact that secondary qualities have their own 'geometries' - form their own a priori systems - does not upset the distinction we are acknowledging. The a priori in secondary quality systems concerns how things look (sound, etc); the a priori in the geometry of space concerns how things are.¹⁵

But what is the status of this concession? Surely very uncertain: it falls into the same rather unprocessed position as that occupied by the acknowledgement of the existence of common sensibles. That there are common sensibles - if there are - looks as if it ought to be significant; but it is quite unclear what its significance is. Both of these points dangle somewhat: they offer no prospect of improving the explanatory work done by primary qualities; they offer no consideration capable of compromising the argument to the effect that subjects of experience can be constituted without recourse to the primary. To concede these two points of difference between primary and secondary qualities does not, as far as I can see, embroil us in a detraction from the

metaphysical autonomy of the secondary.

To say that the squareness of a table is essentially experiential (must be capable of appearing square) is to say that it is a secondary quality. So here we have a quality which is both primary and secondary, as I have set up these notions. It is both capable of figuring in physical theory, possibly at a quite fundamental level, and is essentially experiential. This raises a terminological problem: I am not sure whether it is better to keep the primary status of squareness as instantiated by tables, and take on board the catholicity which will then invade the domain of the primary, or to remove such qualities from the primary and put them exclusively into the secondary category, thereby possibly according some physico-explanatory power to the secondary. I slightly incline to this latter option, because although the table's squareness may figure in a physical theory, there is no guarantee that it will figure in the best such theory, as there is a (logical) guarantee that the posits of the best such theory will. So the status of the table's squareness as a primary quality is sub judice; but its secondary status is beyond question. We must anyway accord psychologico-explanatory power to the secondary; so the extension to according physico-explanatory power would be a matter of degree not of kind. If we are interested in delineating metaphysical natural kinds, the natural division looks like the one between qualities which must figure in a physical science but are not essentially experiential, and qualities that are essentially experiential but need not figure in a physical science. It would surely be sensible to fit the terminology to this division. If we follow this

suggestion, and identify primary qualities with those properties posited by the best physical theory, we do in some sort preserve a connection between the primary and the objective. Such qualities are not themselves experiential; but they are essentially detectable by subjects of experience. This move entails divesting the traditional category of the primary of spatial properties of manifest objects, and putting them into the secondary category.

What now of the purported connection between independence and explanatory adequacy? If we want here qualities which 'will always pull their weight in all explanatorily adequate theories of the world', and when we mean by 'world' the object of fundamental scientific investigation, then it is clear that primary qualities are by definition fitted to service such a project. When, on the other hand, we understand by 'world' the manifest world (the experienced world), then it is clear that only secondary qualities will meet the adequacy requirement (assuming the terminological change canvassed above). For what else in the manifest world, other than the squareness and redness of the table could possibly explain our experience of, respectively, its squareness and redness? So on this reading of 'world' we find that independence and explanatory adequacy come adrift. Both primary and secondary qualities can be explanatorily adequate, depending on which world we are operating in.

Suppose we try to run the explanatory theorizing between scientific and manifest worlds, rather than within each world. Primary qualities of course would be the items carrying the explanatory weight in such a theory: can we vindicate the

implication from explanatory adequacy to independence here? My feeling is against the suggestion: the relation between the worlds is one of supervenience, and I do not think supervenience is a relation the obtaining of which explains (makes illuminating sense of) the supervening level in terms of the supervened-on level. We have already noted (1.5) one important respect in which supervenience inhibits comprehensibility (cf. also chapter 1, note 25). Certainly there is no straightforward implication to independence from explanatory adequacy. We have said enough, however, to show that a fundamental physical theory could not be an absolute conception, or a view from nowhere. This is yielded by the requirement on posited phenomena of detectability by subjects of experience. A theory of the world would have to be capable of being constructed by a subject of thought for it to count as a theory; and the properties which were discerned by that theory would have to be capable of detection by that subject, who would therefore have also to be a subject of experience; so that even if the properties in question were experience-independent, they would still enjoy an essential connection with properties which were not experience-independent, and an essential connection with an experience itself incapable of incorporation into the scientific image. The objective would be the handmaiden of the subjective. So the manifest world is in a clear sense prior to the scientific world. The manifest world necessarily resists reduction to the scientific world. (But it is unclear that the reduction would not work in the other direction.)

There is no absolute conception; there is no view from

nowhere. Crudely speaking, the reason for this is as you would naively expect: an absolute conception wouldn't be a conception; a view from nowhere wouldn't be a view. The problem is not, as Nagel supposes, that a conception or view cannot embrace its own constitution: that seems to me to be unproblematic. The problem is that a conception cannot write out of itself the possibility of its being conceived, not only by a subject of thought, but also by a subject of experience; and experience resists incorporation into such a conception, at least if our internalist intuitions were correct. In that sense, and with this latter proviso, a conception could not be absolute.

Chapter 3: Agency and Time.

3.1 The Ineliminability of A-time.

I turn now to the other main topic of this investigation: freedom and the nature of the free self. In this and the next chapter I consider two fundamental features of our world: that it is in time and that it is in space. Any world capable of being inhabited by subjects of experience and agents must both be in time, and in some sort of space (we saw in the last chapter that the space of material objects is not the only conceivable kind). I shall argue in this chapter that only the free transcendental agent is capable of constituting the temporality of the world; in chapter 4 I shall argue that the parallel thesis - that only a transcendental subject can constitute the spatiality of the world - fails. Time is transcendental, but space is empirical. This distinction will help us to get clear on the nature of the self in the final chapter. We shall also see that the existence of freedom, and hence the reality of time, depend on the internalist position adopted so far. That will yield us a retroactive strengthening of that internalism.

Let us begin our consideration of time with the distinction between A-series time and B-series time. An event is located in A-series time (or A-time) just if it is identified as past, present or future relative to a moment designated as now. The A-series is the series of past, present and future events. Of course there is not only one such determination, but as many such determinations as

there are moments designable as now. One might say that there is a succession of continuously overlapping A-series, or that there is one continuously moving A-series. This latter way of speaking seems to be preferable, because the movement of the present moment is integral to the way we think about time, when we think about it in A-terms. We think of future events as approaching us, becoming present, and then receding into the past. We also think of this movement as unidirectional. Both these features of A-time will figure crucially in the discussion that follows.

In language, A-facts are represented by the device of tense: A-facts are tensed facts. B-series facts, on the other hand, are tenseless. Two events are related in the B-series just if one of them is earlier than or simultaneous with the other. The B-series is generated by the imposition of a dating system on a stretch of history. For the numerical relations between the items of the system, when interpreted temporally, produce B-relations between the events of the stretch of history in question. McTaggart was the first to make explicit the distinction between the A-series and the B-series, and he also introduced the term 'C-series' to refer to the linear ordering relation drawn on by the B-series.¹

It is natural, following McTaggart, to regard the B-series as the joint upshot of the A-series combined with a C-series; that is to say, as yielded by the imposition of a linear ordering relation on the items (say, events) discerned by the A-series. This view would regard both A-series and C-series as jointly constitutive of a B-series. The A-series is needed to constitute the B-series as a temporal series, and the C-series to order the events initially

characterized only in A-series terms. The A-series on its own is not competent to perform this ordering task because it does not have the resources to distinguish two events both characterized as, say, in the past. The C-series achieves the ordering by imposing a metric on the A-related events. I believe that this natural view is correct, that it has important implications for the way we understand time, the connection between time and the self, and ultimately how we think of man in the world. The ineluctability of A-time is, as we shall see, at the heart of the manifest image of man, as I have internalistically characterized it. The fact that, as I hope to show, the manifest image has at its centre a structure of mutual implications between the ineliminability of A-time, internalism and the freedom of the will gives it its peculiar power and resilience; for whatever one might think of the independent plausibility of internalism and freedom, surely no one could jettison the reality of A-time. But then if the connections hold, the manifest image stands firm.

Attempts have been made to do without the A-series by reducing A-facts to B-facts, but these attempts always miss out an essential element in our understanding of temporality. The attempts generally start off with good thought that A-facts are in some sense mind-dependent. It is surely right to say that there is a deep connection between tense and consciousness, and I shall later try to say something about what the connection is. It is not an easy matter getting the connection right, and we must avoid trying to make it too directly or too quickly. Grünbaum is at fault in this respect when he defines the nowness of an event as its simultaneity with a

now present experience of which the experiencer is aware.² For of course there could be moments which are present when all experiencers are (say) asleep. Such moments could enjoy presentness by being part of an A-series, so that there would be previous occasions on which they could be thought of as future, and subsequent occasions on which they could be thought of as past; but it is enough that such a series should exist: it is not necessary that any particular moment in it - or indeed any moment at all - should actually at any time be the object of an experiencer's cognition. It is of course open to Grünbaum to weaken his definition: he may say that a moment is now just if there could be a now-present experience which it is available to be simultaneous with. Such a weakening would certainly preserve the truth of the definition, but I do not think it draws the connection between A-facts and mind in the deepest place. I shall argue that the connection should be drawn with agency rather than experience.

Granted that there is an acceptable sense in which A-facts are mind-dependent - although we have yet to say what that sense is - the next step of the reductionists' strategy should be examined. It is claimed that the A-series, being mind-dependent, is not in rebus because it is always possible to capture A-facts in B-terms.³ The B-surrogates of A-facts are their tenseless truth-conditions. Thus,

(T) It is now the case that p
is true if and only if 'p' is true on the occasion of tokening, i.e. of utterance, inscription or thought that (it is now the case that) p. The truth-conditions of (T) are specified for occasions when 'p'

is not tokened by adding the subjunctive condition that (T) would be true if and only if, were 'p' to be tokened at some time, 'p' would be true at that time. In general, given the indexical character of tensed temporal indicators - given, for example, that tokens of 'now' refer to the moment at which they are tokened (and similarly for other temporal indexicals) - it is always possible to specify tenseless truth-conditions of tensed sentences. This further step in the reductionists' argument is also correct. A-facts do indeed have B-surrogates.

The argument is strengthened by the connection which should be drawn between the tenseless conditions of application of tensed temporal indicators and understanding. To understand a temporal indexical-type I need to know its general conditions of application - what Kaplan called its character;⁴ further, to understand such an indexical on any occasion of use I need to know not only its character but also its Fregean sense.⁵ For example, to understand the word-type 'now' I need to know that it in general refers to the moment at which it is tokened; to understand a token of 'now' (to understand the word-type on some occasion of use) I need to know which time is contemporaneous with the tokening of 'now'. Both character and Fregean sense can be specified tenselessly: the character of the type 'now' consists in its referring to the moment at which it is tokened; the sense of any particular token of 'now' consists in its referring to whichever moment is contemporaneous with the tokening. Neither specification explicitly presupposes A-series time, since neither relies on the movement of the A series, or on its anisotropy as between past and future. I shall indeed

argue that the B-series does presuppose the A-series, but this presupposition cannot be extracted merely from the requirements of understanding.

The line of argument we are considering now exploits the immediate irrelevance of A-series time to understanding as follows: if the tensed fact of a moment's nowness is not reflected in what an understander needs to know in order to understand the temporal segment of the language, how can the moment's nowness be an objective fact? For how can the world transcend language? This is a familiar line of thought, and it is not damaged by pointing out that a speaker needs to know that the Fregean sense of tensed temporal indicators undergoes change from one moment to another. This thought is of course correct. A speaker needs to be able to keep track of thoughts had at other times, and so he needs to know, as Frege said,⁶ how to use different indexicals at different times. For example, since the referent of 'now' varies from one occasion of use to another, a speaker cannot capture at t_1 what he expressed at t_0 when he uttered 'It's F now' unless he now, at t_1 , changes the indexical of his utterance to 'then', and says 'It was F then'. But this explanation of the speaker's ability to manipulate indexicals whose Fregean sense is variable from one occasion of use to another has proceeded entirely in tenseless terms. Change of Fregean sense is not as such an A-series phenomenon, because change is not as such an A-series phenomenon. Change is just the possession of different properties at different times - which is a B-series phenomenon. So far as understanding goes, change in Fregean sense amounts to no more than tenseless variation of that sense from one occasion of use

to another. It does not amount to the movement of the present moment along a B-series. In effect, change does not amount to becoming.

The specification of what a speaker needs to know in order to understand the tensed segment of a language can be given entirely in tenseless terms. But not only can, also must. No specification of the truth-conditions of a tensed utterance which just re-iterated the tense-indicators of that utterance could possibly be correct, because it could not secure the stability of meaning over time. To say that the utterance of 'It's F now' is true iff it's F now, is simply incorrect as a general statement of the truth-conditions of the utterance. Perhaps it is F at the moment of utterance, but at no other time. In effect, A-series specifications of truth-conditions are just specifications of the sense of tensed sentences: they ignore character. But to ignore character is to suppose that languages are not stable over time (that meaning is instantaneous). For A-series specifications can only yield the instantaneous truth-conditions of an utterance or inscription. The specification, therefore, of what a speaker needs to know in order to understand tensed language can and must be given entirely in tenseless terms. The explanation in B-terms of what a speaker needs to know contains the essential core of the dynamics of temporal indexicals. So if there are A-facts in rebus, it will be right to say that they are language-transcendent. The reductionists' conclusion - that there are no A-facts in rebus - will follow if the possibility of such language-transcendence has to be ruled out, presumably on grounds of incoherence. This I do not believe to be an option for us, but it is instructive to see how far the reductionists' strategy can take us:

it takes us to the point where we must acknowledge that the objectivity of the A-series, if such there be, is transcendental with respect to language.

Transcendental with respect to language, but not with respect to mind; this is the place where we must call a halt to the reductionists' advance. Prior has convincingly argued that there is no tenseless surrogate for the feeling of relief expressed in an utterance of 'thank goodness that's over'.⁷ Of course there is a tenseless specification of the assertibility-conditions of such an utterance: it is assertible just in case the unpleasant event alluded to, E, is earlier than the utterance, U. But the tenseless specification cannot reach as far as the content of the mental state of relief, which is to the effect that E is over, not that it is earlier than U. The tenseless specification does not, and cannot, explain why it is appropriate to feel relief after the event E. For the temporal relation between E and U stands eternally. If relief were directed onto the mere fact that E is earlier than U, there would be no reason why it should be felt at the time of U rather than of E. In fact of course the content of the relief is such that it can only be rationally felt after E: it is a feeling of relief that E is over; there is simply no B-surrogate available to specify the content of this relief.

Tense is transcendental with respect to language but not with respect to mind; we can add some detail to the emerging picture by noting that tense is also transcendental with respect to rationalization for action. In explaining why I perform an action (at a particular time), one needs to cite factors (centrally,

beliefs and desires) which are motivationally significant (at that time). This explaining can and must proceed entirely tenselessly: particular times (such as I can have A-attitudes towards) are not motivationally significant. What are significant are empirical features of myself and the world around me - features which are arrayed tenselessly.

This principle - that particular times cannot have motivational significance - is the counterpart in psychology of the scientific principle that particular times cannot have physical significance. It might be thought that the following kind of scenario provided a counterexample to this: I leap up at 3 p.m. to go to a meeting, remembering suddenly that there is a meeting at that time. Here it might be said that my action is jointly explained by my knowledge that I have a meeting at 3 p.m., my desire to attend it, and my realization, at 3 p.m., that 3 p.m. is now. The latter realization is certainly a mental state of the sort which, following Prior, we agreed has no tenseless surrogate. But it is not the case that this component interlocks motivationally with the action, thereby comprising an ineliminable feature of its rationalization. It is not the nowness of 3 p.m. that explains my leaping up at 3 p.m. to go to the meeting. Rather, it is the fact that 3 p.m. is the time of the meeting that explains my going the meeting at 3 p.m. And it is my forgetfulness of the meeting which explains the suddenness of my manoeuvres just before 3 p.m. My realization that 3 p.m. is now does not itself do any explanatory work.⁸

The argument for the eliminability of tense cannot then be met by claiming that tense enters into truth-conditions or

assertibility-conditions, and hence enters into a specification of what a speaker needs to know in order to command a language. And it cannot be met by locating tense in rationalization for action. But we have seen that tense can nevertheless enter ineliminably into the content of some mental states, although the occurrence is transcendental with respect to language and rationalization for action. How should we register this occurrence? I believe we must reply to the reductionists by appealing to the point already touched upon, and claim that tense is ineliminable because it is needed to constitute the B series as a temporal series. Tenseless specifications of relations between events do not eo ipso order them in time unless there is warrant for interpreting the ordering relation employed as a temporal relation. That warrant is supplied by the A-series.⁹ Those who have argued for the eliminability of the A-series in favour of the B-series have in effect just helped themselves to a temporal interpretation of the ordering relation employed by the B-series. But in advance of an integration of a B-series with the A-series, there is no justification for regarding the series in question as amounting to anything more than a C-series. To interpret the ordering relation of such a C-series, and hence to constitute the series in question as a B-series, we need to know that it is fully integrated with the A-series. For an instant at t_0 to count as earlier than an instant at t_1 it is necessary that t_0 is past from the point of view of t_1 , and that t_1 is future from the point of view of t_0 . That in turn requires that the instants in question be available to be inhabited by minds equipped with the concept of tense, or, in other words, with a temporal perspective.

To enjoy a temporal perspective must in turn require the inhabiting of a moving now. Movement of the present moment (one of the two definitive features of A-time mentioned at the beginning) must be understood if the formulation ' t_0 is past from the point of view of t_1 ' is not itself to be a B-formulation, in effect just a trivial variant of ' t_0 is earlier than t_1 '.¹⁰

Connecting it with a moving temporal perspective respects the mind-dependence of the A-series, without relinquishing the objectivity of tensed relations. It ranges tense along with other phenomena, such as colour, acknowledged to be both subjective (in the sense of being essentially experiential) and objective (in the sense of being genuine components of the experienced world). And as with these other phenomena, the mind-dependence of tense is not correctly expressed by projectivist locutions. To say that the mind projects the A-series is to say that we can form some conception of what the screen is like - what the world is like in advance of the act of projection. But we can form no such notion. The four-dimensional block universe does not provide us with a candidate for such a screen, because it would then be wholly mysterious why we temporalized one dimension rather than another, or indeed why we temporalized any dimension at all. If we had a choice in this matter, there would be no principle on which to exercise it. But in fact it is absurd to suppose that we have such a choice. The idea of the block universe arises when we spatialize the temporal dimension of the actual universe - or, more accurately, it arises when we suppress the specifically temporal component of the time dimension (the A-series), leaving only the spatial or quasi-spatial C-series.

It does not provide us with an idea of what the world in itself is like, in advance of some mental act of temporalization. Furthermore, there is no sense to the notion of consciousness as a whole confronting and temporalizing a block universe. For consciousness is in time, but the block universe is outside time; so such an interaction would not be possible.

This line of argument also disposes of the alternative hypothesis that individual consciousnesses (rather than consciousness as a whole) confront the block universe in order to project temporality onto it. The contention here is graphically conveyed in the picture of individual consciousnesses crawling up their respective world-lines. But the picture is incoherent. For such crawling consciousnesses to be conscious, they would have to interact with the world: they would have to perceive things and perform actions in the world. Centres of consciousness just are subjects of experience and agents. But no such interaction could occur between a consciousness and a timeless world. When something is perceived by a subject of experience, or manipulated by an agent, it is taken up into the time-order of the consciousness in question; it ceases to be a now-thick slice of a space-time worm and becomes an object existing in time, that is to say an object with spatial but not temporal parts.

It has been suggested¹¹ that there is no extra difficulty in explaining how tensed experience could arise from tenselessly located events beyond the difficulty of explaining how, say, colour experience could arise from phenomena that lack colour, or non-spatial mental events from spatially disposed neural events. In

the light what I have already said, I think we can agree that there is indeed a deep problem about the provenance of these sorts of subjectivity, but I think the above considerations are sufficient to show that there is an extra problem about time. The difficulty we encounter when we envisage tenseless events giving rise to tensed experience is not just that there is a disparity between the status of the cause and that of the effect, as it is with the other subjective phenomena. More fundamentally, the difficulty lies in the availability of the causal relation itself. For events interact causally in time, so that if tensed experience has causal antecedents at all, the tensed nature of the experience will rub off on those antecedents. If it is right to say that tensed time is mind-dependent in that the possibility of mindedness is what secures the possibility of time - just as the possibility of experience is what secures the possibility of colour - still there is even less reason than there is in the case of colour for saying that tensed time is therefore not objective.

3.2 The constitution of A-time: the failure of causality.

If it takes an A-series to constitute the temporality of the B-series, and hence to constitute time, what does it take to constitute the A-series? So far we have only the general and in itself rather unspecific answer that the A-series is mind-dependent: what aspect or function of mind is the A-series dependent on? I want to argue that only one aspect of mind is relevant to the constitution of the A-series, namely agency; and I shall also argue that the sort of agency needed to perform the constituting role is free agency, in the full incompatibilist sense of 'free'.

As I have stated, the A-series supplies two fundamental features of time: its movement and its directedness, or anisotropy. Time is anisotropic in the sense that there is a genuine distinction between past and future which applies in one and the same orientation to all temporally located beings. For each now, there is just one temporal point of view which all such beings share. The future and the past are not structurally isomorphic. In this respect time differs from space: for while spatial dimensions are inherently ordered, they are not directed - no spatial point of view is privileged. Any candidate for the constituting role of A-time must secure its anisotropy if it is to have a chance of success.

It might be thought, and it has often been argued, that causality can supply the necessary directedness. Causality is

certainly an asymmetrical relation (in the logician's sense); the question is whether its asymmetry is prior to, and hence potentially supportive of, temporal asymmetry, or whether the dependence is the other way round. The matter requires some delicacy. The first point to make is that asymmetry is not on its own enough to constitute anisotropy. To say that causation is an asymmetric relation is not yet to say that causal considerations can yield a particular and unique direction to a stretch of history. But for a relation to constitute the A-series we need the following to be the case: if we select an instant in a stretch of so far undirected history and label it 'now', our candidate relation must yield a unique future and past. It must not be the case that we could equally well apply the relation in the opposite direction. That would mean that there was no unique temporal point of view associated with now: it would license the adoption of either of two alternative and mutually incompatible temporal points of view, according to the first of which events in one direction were future and in the other past, and according to the second of which events in the other direction were past and in the one direction future. In effect time would be assimilated to a dimension of space, for while each spatial dimension is essentially ordered, neither of the two available orientations associated with that dimension is privileged. Take any two points X and Y situated along such a dimension. From some point of view X is to the left of Y; but there is available another equally good point of view from which Y is to the left of X. The road from London to Edinburgh is equally the road from Edinburgh to London. Spatial dimensions are ordered but isotropic; a candidate

for the role of constituting the A-series must secure anisotropy: to repeat, if from some point of view taken as now a moment X is discerned by the relation as future and a moment Y as past, there must not be available an alternative equally good superimposition of the relation according to which Y is future and X is past. That a relation is asymmetric is not good enough for anisotropy: the relation 'to the left of' is asymmetric, but it is not superimposed on space anisotropically. I shall argue in this section that causality cannot supply the uniqueness of time's direction.

It has been suggested by Michael Dummett¹² that we should separate the question of the reversibility of causation from that of the reversibility of explanation; he has argued that physical processes are not reversible without loss of explanatory intelligibility, but that this fact does not tell us which direction causation runs in. But I do not think it can be correct to separate out two questions here. Strawson has made familiar the thesis that there is a tight triangle between the ideas of causation, the exercise of causal dispositions (upon which explanation depends) and the idea of a thing of a certain sort.¹³ The suggestion is that the idea of a thing of a certain sort does not, as Hume thought, come in advance of the idea of that thing possessing certain causal dispositions. And the idea of causation is, at least primarily, the idea of a thing exercising causal dispositions. If this is right, as I believe it is, it follows that the explanation of the way things behave in the world (which must appeal to their sorts) cannot be given independently of the direction of causation.

Suppose, for the sake of illustration, that a large boulder

careers down a hillside and crushes a hut, upon the fragments of which it comes to rest. An explanation of the crushed state of the hut will advert to its fragility, and its situation in the path of the falling boulder. But now it cannot be still up for determination which direction causation acted in. The explanation must appeal, via sorts (such as 'boulder', 'hut', 'wood' etc.), to the causal dispositions of large boulders to crush relatively fragile objects which they hit with sufficient impact. The explanation appeals to the hardness of the boulder and the fragility of the hut: both of these properties are causal dispositions. The validity of the explanation depends on the boulder and the hut having such dispositions, and exercising them on this occasion. But the possession and exercise of these dispositions is compatible with just one causal direction. Were we to suppose that causation ran in the opposite direction, we should have to imagine that very different causal dispositions inhered in the boulder and the hut. Specifically, the hut would not be, in respect of large boulders, fragile, but rather would have the power to propel them up hillsides. And piles of broken matchwood with boulders lying on them would have the ability to spring into huts with such propulsive powers. If causation ran in the opposite direction, so would explanation. (Of course it goes without saying that such a world would not be immediately intelligible to us, accustomed as we are to the causal dispositions in our world; but that is another matter.)

We can simplify our task, therefore, by assimilating the direction of causation and the direction of explanation. Where A and B are events, A causes B iff the occurrence of A explains the

occurrence of B, and not vice versa. This is because, as I have said, causal explanation makes primary reference to the exercise of particular causal dispositions, and particular dispositions can (at least in many cases) only act in one direction. Nothing in this discussion settles the question whether causal processes are reversible - that is what we still have to consider. All that I have claimed is that causal processes cannot run against the direction of explanation. But perhaps they are jointly reversible.

If they were jointly reversible, an object would have two different sets of causal dispositions. The sets would not necessarily be disjoint: tennis balls, for instance, would be bouncy whichever direction causation ran in. But for most and perhaps all objects, at least some causal dispositions would belong to just one of the two sets. Which set of dispositions got activated would depend upon which direction causation ran in, and the question we have to address is whether or not that is an arbitrary matter. But we should note first that there is nothing in the bare metaphysical truth that objects come packaged with causal dispositions to rule out the possibility that more than one set of such dispositions may be available to an object.¹⁴ So far, the direction of causation (and with it that of explanation) is up for determination.

I shall discuss both causation and explanation in order to ascertain whether either can be superimposed irreversibly on a world of objects of certain sorts. Of course, if the above discussion is correct, it is not necessary to consider both cases; but it is interesting and illuminating to do so, and in any case provides an independent check on the previous discussion (of a negative kind: if

the following discussions reach different conclusions, the above claim must be false).

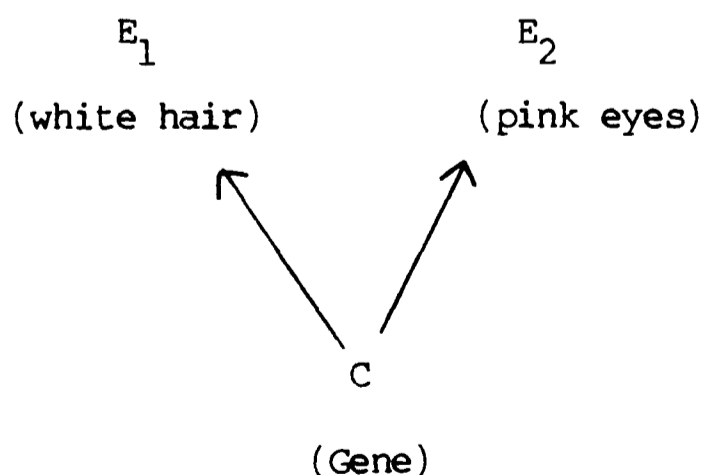
I take first the case of explanation of causal processes. In one of the passages already referred to (footnote 12), Dummett argues that gross events in the world exhibit a unique explanatory asymmetry. He considers the case of an apple falling from a tree and rolling some distance. In reverse the process would look like this: an apple would suddenly start moving along the ground, gathering momentum until it would suddenly jump up onto a branch. According to Dummett, gross observation would find two features of the time-reversed process inexplicable: the instant at which the apple started rolling and the direction of its motion. But of course these two inexplicable facts are precisely mirrored in the forward direction. It is inexplicable to gross observation why the apple falls from the tree at the moment it does, and why upon impact it rolls along the ground in the direction it does. At the level of the manifest world, there are inexplicable features of both forward and reverse scenarios. Explanation must postulate sub-observable features of the situation to account for these features; but it can do so in the reverse direction just as well as in the forward. In the forward direction, we explain why the apple drops from the branch at the moment it does in terms of the ripening process, and the associated weakening of the stalk. In the reverse direction we can explain why the apple starts rolling at the moment it does in terms of the reverse-rottening process: the apple becomes less and less rotten until a moment comes - in principle predictable - at which it must start moving, just as in the forward direction the

stalk weakens until a moment comes - again, in principle predictable - at which the apple must drop.

Another example of a causal process which is said to make more sense in one direction than another is Popper's case of a point source generating waves in a tub of water.¹⁵ The coherence of the wave pattern as it spreads out is explicable in terms of the unitary source, but in the reverse direction the coherence of the incoming waves would look like an extra-ordinary coincidence. I want to accept that this description of the case is correct, but deny that the features of the case which warrant finding an explanatory asymmetry in it between forward and reverse directions can be generalized over just the forward direction. Firstly, there are many causal relations or parts thereof which exhibit no such coherence or coincidence in either direction. But secondly and more importantly, there are many phenomena which exhibit coincidence in the forward direction and coherence in the reverse. A car-crash is a case in point. One can easily envisage possible scenarios where in terms of the manifest phenomena the process makes more sense in reverse: there is a central event from which events radiate coherently outwards; in the forward direction the smash-up might be the most extra-ordinary and unfortunate coincidence. Evolutionary adaptation is another example of a process that would enjoy a coherence looked at in reverse which it lacks when looked at in the forward direction. (In general, processes which could bear a teleological explanation are going to be such that, if the explanatory resources are restricted to efficient causation, they will be more intelligible in reverse.) So while I think we should accept Popper's

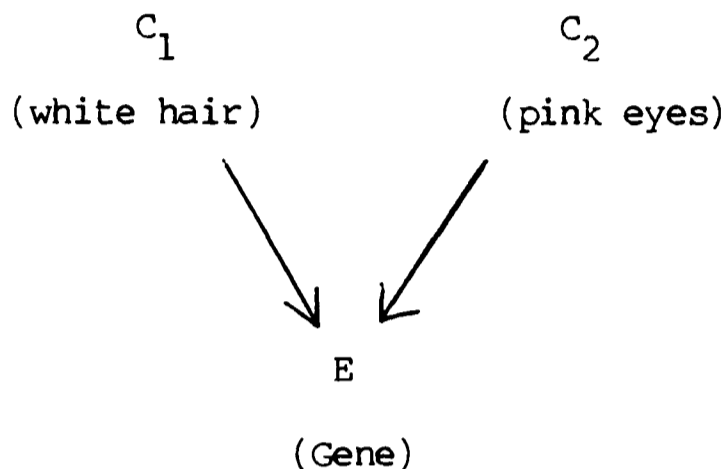
point for his particular example, the example chosen does not share its crucial features with all and only forward processes. Many reverse processes share those features; and many processes do not have them, either in forward or in reverse directions. As far as explanation in rebus is concerned (excluding human behaviour), we ought to say that explanation could operate perfectly well in either direction. Teleology would map onto efficient causation and vice versa, and coincidences with more than one effect would map onto coherences and vice versa.

Let us now turn to an examination of causation itself: is it, when considered in rebus, isotropic? (We can restrict ourselves to the consideration of causation in rebus; causation by human intervention introduces extraneous elements, and if causality can direct time, then that ought to show up in the extra-human causal order.) Two problems confront the claim that causation is isotropic. The first is raised by Mackie: it concerns whether the logical structure of causal sequences involving common effects produced by a joint cause can be preserved if the processes are run in reverse. Consider Mackie's case¹⁶ of a single gene producing white hair and pink eyes. The causal structure looks like this:



Given an appropriate causal field, C is necessary and sufficient for

(E_1 & E_2), and sufficient for each of E_1 and E_2 separately. (C & E_1) is necessary in the circumstances for E_2 , as (C & E_2) is for E_1 . On the other hand, Mackie argues, if we reversed the time direction, so that the picture looked like this:



then C_1 would not be necessary for E (nor C_2 necessary for E), but each of C_1 and C_2 would be sufficient. We cannot tie C_1 and C_2 to co-occurrence by stipulating that E occurs, as we can tie the effects E_1 and E_2 to co-occurrence in the forward direction by stipulating that C occurs. The time-reversed image of the forward process would in fact have the logical structure of a case of overdetermination.

But suppose that, instead of regarding the reverse process as a case of overdetermination, we regard C_1 and C_2 as jointly sufficient and severally necessary for the effect, rather than as severally sufficient and neither jointly nor severally necessary for the effect. If we take this line, then although the logical structure of the cause-effect relation changes under time-reversal, the causal relations between the various items which are the causes and effects remains. In both forward and reverse directions, the

occurrence of the gene is, given the respective causal relations, necessary and sufficient (within the causal field) for the occurrence of the combination of white hair and pink eyes. And the co-occurrence of the gene and white hair (/pink eyes) is jointly necessary and sufficient for the occurrence of pink eyes (/white hair). We do not have to regard the reverse process as a case of overdetermination. In fact it is more plausible, in the present instance, to take white hair and pink eyes as, in the reverse process, jointly rather than severally sufficient for the occurrence of the gene. That preserves the logical structure between gene, white hair and pink eyes. What undergoes transformation is (ex hypothesi) the causal status of these items, so that in the reverse process each cause is no longer sufficient for the effect, whereas in the forward direction the cause is sufficient for each effect. Again, in the reverse process the effect is, given the causes, not necessary for the occurrence of any other effect, whereas in the forward direction each effect is, given the cause, necessary for the occurrence of the other effect.¹⁷

Still, it might be objected, the difference remains that in the reverse process two causes have to be tied to co-occurrence if the effect is to occur, whereas in the forward process two effects have to be tied to co-occurrence if the cause is to occur; and this latter might be felt to be less problematic, since the occurrence of the effects is secured by the prior occurrence of the cause, whereas in the reverse process the effect cannot secure the co-occurrence of the causes, just because it is not prior to them. This is certainly an important difference between the two processes, but it does not

threaten the coherence of reverse causation. It cannot do so, because there are many forward processes which share the feature of the reverse process under consideration. For example, the occurrence of an explosion in certain circumstances requires that (say) a spark and inflammable material be tied to co-occurrence; the explosion itself cannot secure their co-occurrence, because it is not prior to them. In reverse, such an explosion would mimic in its causal relations the logical structure which the gene/white hair/pink eyes causal nexus has in the forward direction. But that does not make the reverse direction more coherent or comprehensible than the forward direction: we are quite used to the phenomenon of causes combining (often accidentally) to produce effects. We must conclude our discussion of Mackie's case as we did with Popper's case: while the particular example chosen looks different under reversal, it nevertheless resembles under reversal many other forward processes. There can be no bar, then, to the possibility of reversing causal sequences which involve separate effects having a common cause.

We have just been considering causality at the manifest level - causal interactions between gross observable objects. Now it is often suggested - and this is the second problem for us - that the irreversibility of certain fundamental processes or laws could supply time with its direction. But in fact such irreversible processes have nothing to do with time's direction. For any law L covering such a process P , there is a time-reversed transform L' covering the inverse process P' under time-reversal.¹⁸ It is no good saying that the fact that P occurs rather than P' underpins time's anisotropy. The fact here appealed to simply presupposes the

direction of time. The question is what entitles us to regard the obtaining of P rather than the obtaining of P' as the fact of the matter. The only possible answer to this question is that the time-direction, antecedently fixed, determines which of P and P' obtains. The bare examination of the processes and their accompanying laws cannot tell us this.

These objections apply quite generally to the suggestion that physical processes or laws could provide time with its direction. But there is also a separate and specific response to the contention that fundamental such laws or processes might do the trick. It is that the descent from the manifest to the scientific level nullifies any possibility of success here. That is because the connection between the scientific and manifest worlds is at best contingent. But time is an necessary feature of the manifest world, so that anything which is competent to constitute the time-direction must be able to do so in the manifest world, however things are at the fundamental physical level. If we rested the time-direction on some feature of that level, we would have failed to secure its necessity.¹⁹

My conclusion from this discussion is that causation in the extra-mental manifest world, while being asymmetric, does not in itself impose a unique direction on the succession of events in the world. And the explanatory status of causal relations would not be damaged by time-reversal. Causality is not an anisotropic phenomenon, and so is not suitable to establish the unique direction of the A-series. If from one point of view A is construed as causing B, there is available another equally good causal point of view from

which B can be construed as causing A. Causality does not secure us a unique future and a unique past.

We should not confuse the question of the direction of causation with the problem of whether backwards causation is possible. Backwards causation is causation which runs against the time-direction. I am not claiming that this possibility is coherent; indeed I think that it is definitely not coherent, for reasons that I shall give later. I would certainly want to say that if the direction of any causation is fixed, then the direction of all of it is fixed. What I claim is that the direction of causation - any and all of it - is not fixed in advance of the time-direction, and so is not antecedently available to constitute the time-direction. If this is correct, the fixing relation must operate in the opposite sense: it must be the direction of time which fixes the direction of causation, and not vice versa. But it remains the case (as I shall argue) that the time-direction and the direction of causation necessarily coincide.

3.3 The constitution of A-time: the need for free agency.

We are still looking for something to constitute the unique direction which time has. What is required of a candidate is that it supply anisotropy, and hence yield a genuine distinction between future and past: given an instant designated as now, it must secure a unique orientation such that there does not exist an alternative point of view from which the opposite orientation is available. The ordinary view of what constitutes time's anisotropy is surely this: the difference between the future and the past is that the past is fixed (we can do nothing about it) whereas the future is open (we can affect it - not in the sense of changing it, but in the sense of making a difference to it). This characterization is surely correct as far as it goes; the question is how much metaphysical bite it should be given.²⁰

It should be noted that on a very natural interpretation the ordinary view yields and connects two different asymmetries - one concerning action and one concerning knowledge. The past is fixed both in the sense that I can now do nothing about it, and in the sense that I can get to to know how it went. The future is open both in the sense that I can affect it, and in the sense that I cannot now know how it will go. The two asymmetries necessarily fit together in just the above way. I cannot affect what is epistemically available to me; I cannot know about what I can

affect. If it were possible to affect the past, our belief in the knowability of the past would thereby be undermined; similarly, if it were possible to know about the future, our belief that we can affect it would be undermined. (The epistemic asymmetry incidentally guarantees that we all inhabit the same time-order: I cannot be free to affect what you know about and is therefore fixed.)

The phenomenon of intention does not affect the epistemic asymmetry: intention cannot amount to knowledge, in the relevant sense, of my future actions, because it is always possible for the execution of an intention to be frustrated, perhaps by my changing my mind. (I can of course decide not to change my mind; but I may later renege on that decision.) The only point here which matters for my purposes is the epistemic difference between memory and intention; if someone merely wishes to call some cases of intention cases of knowledge, I have no argument with him provided the inferior epistemic status of intention be recognized. If one has a memory, it follows that he knows about the past; on the other hand it does not follow that if one has an intention, he knows about the future. It may turn out subsequently that his intention did amount to - if you insist - knowledge, but the certification of his mental state as an intention does not depend on that success. This is not a verbal point: it is not just that we only call apparent memories 'memories' if they amount to knowledge, but place no parallel restriction on intention. Rather, consider that there could be no such thing as an apparent intention: if I fail to implement an intention, that does not go to show that I did not have the intention. States of intention are not aiming to amount to knowledge

of the future; they are aiming to shape it. States of attempted memory, on the other hand, are aiming to achieve pure cognitive status: they are therefore liable to a failure to which states of intention are not liable. The difference in purpose is reflected again in difference of epistemic position: I can have an intention (and know that I have an intention) without thereby acquiring cognitive access to the future; I cannot have a memory (let alone know that I have a memory) without thereby acquiring such access to the past.

I take the above characterization of the difference between past and future - connecting an action-asymmetry with an epistemic asymmetry - to be the main thrust of the ordinary conception of that difference. What is its full metaphysical force? Suppose one asked: could I not know of some action of yours, which from your point of view has yet to be performed, and is the sort of action which you can regard as yours, that is to say an action not performed under any kind of duress? And if so - the question urgently presses - could not the generalized instantiation of this type of scenario (the knowledge of future actions which the agents could still regard as theirs) be capable of sustaining a time-direction? What this question asks us to consider is whether compatibilist agency might be able to constitute the time-direction, even if it were unable to employ a notion of epistemic asymmetry as between past and future, allowing that the future as well as the past may be knowable (because fixed). It tries to secure the result that acting into a fixed future could nevertheless be genuine agency, by arguing that if the provenance of the agent's actions is suitably grounded in

beliefs and desires which the agent can regard as his own, then he can intelligibly appropriate to himself the movements that arise out of these beliefs and desires. Whether or not such appropriation is imaginatively possible is not my concern here; rather, I am interested in whether such agency could yield the anisotropy of time. If we attempted to retain the past-future asymmetry while letting go of the epistemic asymmetry, we should have to follow Russell and Ayer²¹ in regarding it as a mere accident (logically a mere accident) that we are equipped with the power of memory but not of precognition. On the ordinary view, since the epistemic asymmetry is bound up with the anisotropy of time, and hence with time itself, our lack of a faculty of precognition and possession of a faculty of memory is no accident or contingency, but a deep and necessary truth.

What, on a compatibilist view of agency, could constitute the asymmetry between past and future? It is very hard to see how anyone could ditch the epistemic asymmetry and still retain a notion of anisotropic time. But perhaps we have been too hasty in supposing that the compatibilist cannot entitle himself to the epistemic asymmetry. Suppose first it is objected against the compatibilist that on his picture my actions are in principle predictable - and perhaps even precognizable - because fixed in advance. So long as the way my actions get fixed is intelligible (along causal and/or rational lines), there seems to be no particular difficulty in this claim. But then, the objector continues, the epistemic asymmetry between past and future has been done away with, at least in principle, and so therefore (given the plausible assumption that the

epistemic asymmetry is partly constitutive of the time-direction) has time itself. Unfortunately, however, this objection fails to take account of the well-known impossibility - on any account of the matter - of my actions being predicted or precognized by me. This impossibility arises because the agent's purported knowledge of his future actions could not be sealed off and prevented from inputting into the causal or rational grounds for the action itself, and hence possibly changing it: I cannot know about my future actions because any purported mental state of knowledge could frustrate itself by altering or preventing the action purportedly known to be going to occur. If I then try to accommodate the effect of this 'knowledge', I give myself the problem of accommodating the effect of the accommodation. And so on. Very well, the objector might say, but aside from this epistemological loophole (as we might call it), there would be no bar, on the compatibilist picture, to the agent thinking of his actions as in principle knowable in advance, although knowable only by someone else. As far as that consideration goes, the objector claims, there is no epistemic asymmetry in sight. But perhaps, it might be replied, the epistemological loophole - which certainly yields some sort of asymmetry - yields enough of an asymmetry to set up the time-direction.

Before tackling this question, let us pause briefly to contemplate the picture of the agent which the compatibilist is offering us (we shall need it later in the argument). What emerges when we relinquish the component of freedom which the incompatibilist wishes to insert into the agent is a picture of what one might call a 'specator-agent': an 'agent' who regards his

actions as - at some level - things that happen to him rather than things he initiates himself. This description of the spectator-agent seems to need a gloss, if only because the compatibilist will protest that he does indeed regard actions which we would ordinarily call 'free' as things the agent does rather than things that merely happen to him. The compatibilist precisely wants to say that an action can be both one's own (performed as a matter of one's own choice) and causally or in some other way determined (in the sense that one could not have acted otherwise). Now the point of the above insertion of the rider 'at some level' into my formulation of the compatibilist thesis was to register the protest I have just expressed, while at the same time holding on to the fact that there is something along the lines of metaphysical freedom which the compatibilist normally wishes to deny to the agent and which the incompatibilist normally wishes to accord him. Exactly how we characterize this extra nescioquid, which turns what the compatibilist calls freedom into what the incompatibilist calls freedom, is not to the point here. I just require some dummy formulation to do the work. (We can at least know this much about the nescioquid: it is a bar to predictability; and it supplies the possibility that I could have acted otherwise, in the strong sense that the world could have been exactly the same up to the moment of decision, at which point I could have chosen differently.)

But I can specify a bit more precisely what the content of the above formulation is. Part of what the incompatibilist is driving at when he complains that his opponent's picture omits genuine freedom is this: it is impossible to fit deliberation into a picture which

incorporates general determinism of some sort running from past to future. I cannot rationally deliberate about what I believe to be fixed; I cannot deliberate about what I know to be fixed, whether the knowledge of what I shall do is in principle available to me or not.²² This claim might be countered. It might be said that I can deliberate in the full knowledge that the outcome of my deliberation is fixed in advance. On this view the agent still has to decide what to do in any situation; he is not released from decision-making even if he knows that - perhaps at a very deep level - he is not genuinely free. I do not believe this claim is coherent; at the very least it is surely irrational for an agent to deliberate if he knows or believes that he is not free.²³ If it is irrational for him to deliberate, then he ought not to do so: he should sit back and allow the sweep of events to carry him along. To do this would not be to succumb to fatalism; it is the merest common-sense in the face of the knowledge or belief that one's deliberation could not affect the outcome.

Of course one might believe that one's deliberation would affect the outcome, in the sense of making a difference to it. Could it then be rational to engage in it? Let us suppose that the agent believes that there is likely to be a better outcome if he deliberates; he is now wondering whether to do so. Well, either he believes this decision is free or he doesn't. If he is a compatibilist, he is likely to believe that at some level this decision is not free; at least that is the case we are interested in. In that event the above claim still applies: if he believes his decision is not free, it is not rational for him to attempt to

influence it one way or the other. I am just trying to clarify the point that on the compatibilist's picture of agency, there must be a sense or a level at which the agent either cannot, or (more weakly) ought not to engage in deliberation. If he ought not to deliberate, then - in the place where he ought not to do so - the agent ought to conceive of himself as being swept along by events. And if he ought to conceive of himself in this way, it follows that he can do so. For the compatibilist, the hypothesis of an agent who does so conceive himself (that is, the hypothesis of the spectator-agent) must be coherent.

However difficult it may be for us to imagine being such spectator-agents - since I take it that our ordinary intuitions are strongly incompatibilist, at least as regards our own actions - the coherence of the hypothesis that there could be (let alone that we are) such beings is therefore required by the compatibilist. What is required is the coherence of the suggestion that there could be subjects of representational experience, such that temporal precedence could be part of the content of that experience - the subjects would witness a passing show - but also such that the subjects would not conceive of themselves as free agents, in the sense of the above formulation (with gloss): they would not conceive of themselves as having deliberative powers to affect the future. They could indeed appropriate some of their bodily movements to themselves, regarding them as their actions, but they would think in a compatibilist way about these actions, thinking of them as fixed in advance, as not now negotiable. The question is: could such 'agents' see their actions as effecting a distinction between past

and future, and so as constituting the time-direction?

We can now return from the digression to the question of epistemic asymmetry. At the point where we left the argument, we had agreed that the compatibilist agent might, given the epistemological loophole, be capable of directing his time-order, since he cannot himself think of his past and future as epistemically on a par. But, our objector might now respond, surely observers of him can. If it is possible in principle for an outside observer to know as much about the subject's future as about his past, it might be said that this discrepancy between the subject's attitude to his existence in time and the observer's attitude to it is intolerable. For surely either the epistemic asymmetry is there or it isn't. If it is not available to outside observers, then it would surely be unacceptable to allow the subject's view of himself to have priority. He must think of himself as others think of him, as located at a present moment such that there is no essential difference between his past and his future. In spite of the epistemological loophole, he must regard the public perspective as alighting on the real facts of the matter, so that the loophole yields at best a subjective illusion - a necessary illusion, no doubt, but an illusion nonetheless. Or so the objection goes.

But I do not think this argument is successful. Even if the epistemic asymmetry forced on each agent is an illusion, as indicated by the fact that the public view on his actions discerns no such asymmetry, still the mere fact of illusion in each case is enough for the ability of each agent to direct his time. The future for each agent is just that stretch of time which that agent cannot

know about with respect to his own actions, and the past that stretch of time which is available to be known about (irrespective of topic). The epistemological loophole is sufficient to secure individual epistemic asymmetries, and hence is sufficient to permit each individual to direct his own time. So compatibilist agency apparently can construct individual anisotropies; can it construct a general, intersubjective anisotropy? Can it guarantee that all agents inhabit exactly the same time-order?

According to the objector compatibilism allows that there are conditions under which I may in principle know how you will act, even if I cannot know how I myself will act. But in fact this objection seems misplaced, because the conditions under which I can supposedly know how you will act must include the provision that I do not tell you of my cognition, or that it does not seep through to you in some other way. But I cannot know now how I will act; so (at least) I cannot know that I will not tell you of my prediction concerning your action. The loophole accordingly destroys not only my chances of knowing how I am going to act; it also infects my chances of knowing how you are going to act. In principle I can, as one might put it, near-as-damn-it know how you will act, because I can all but know that my prediction will be sealed off from you: I can form the strong intention not to communicate my prediction to you, and take precautions in other ways. But I cannot know tout court what you will do, because I cannot know tout court what I will do: 'near-as-damn-it' knowledge is not knowledge. It follows that, for all that has been adduced so far, the compatibilist can construct a general epistemic asymmetry from the sundry individual

asymmetries which the epistemological loophole supplies him with; hence he can, it would so far seem, construct general time.

The objections to compatibilist agency raised so far accordingly fail. But we can use some of the bits of the wreckage to mount an attack which works. We start by taking a critical look at the kind of ignorance of the future which is appealed to by the compatibilist in his construction of a general epistemic asymmetry. For there is a crucial difference between the compatibilist's conception of ignorance of the future and the incompatibilist's conception. The compatibilist's ignorance is not simply blank ignorance: it is, as we have seen, near-as-damn-it knowledge. One finds oneself wanting to protest that if the epistemological loophole is adequate to yield an epistemic asymmetry between past and future, nevertheless the asymmetry yielded is purely adventitious: it hovers above a time-order in which the future is really just as fixed as the past. It does not seem to engage in the right way with the necessary anisotropy of time. One might try to give weight to this objection by arguing that the loophole is a purely contingent feature of compatibilist agency. If this argument could be made to work, it would destroy the compatibilist's ability to direct time, since he would have failed to guarantee that one of the necessary conditions of time (namely epistemic asymmetry) was a necessary condition of agency as he conceives it: only if the epistemological loophole is a necessary feature of compatibilist agency is it competent to be the resource out of which the necessary directedness of time is constructed.

It turns out, I believe, that the epistemological loophole is

in fact a contingent feature of compatibilist agency. (Here is where we need the digression on the spectator-agent.) The argument for this runs as follows. It cannot be ruled out that there might be fixed points for epistemic input to the subject's cognitive states.²⁴ A fixed point of a function is a value in the domain of the function which is mapped by the function onto the same value in the range. Now it ought in principle to be possible for outside observers of a spectator-agent to calculate the fixed points for a number of knowledge-input/behaviour-output functions relating to that agent, and then simply feed him with predictions that he will perform the actions which comprise those fixed points. Given that the actions are fixed points, the agent will perform them, either in spite of or (partly) because of hearing the prediction. This spectator-agent's anterior beliefs that he will perform the actions in question do not go towards frustrating those beliefs: his beliefs about his own future actions approximate to the status of knowledge even more closely than the near-as-damn-it knowledge which he has in principle in respect of the future actions of others; it makes no difference either if the spectator-agent is informed that the predicted actions are fixed points (if he realizes that he is being manipulated): that has been allowed for in the manipulators' calculations.

In principle, the following scenario would not be excludable under compatibilist considerations: we are all spectator-agents, the universe is finite, and we are able to complete a scientific description of it (including our own interaction with the universe). If we, or our descendants, found ourselves in this position, it

would in principle be possible for us to calculate the fixed points for a range of knowledge-input/behaviour-output functions, and then simply feed ourselves with these fixed points. Given that the calculations would take into account the future states of the world as well the future behaviour of the agents in it, it would be possible for us (or our descendants) to acquire a complete knowledge of the course of the future. Or, at least, it would be possible to acquire such knowledge modulo the absence of anything like divine intervention. In such a scenario, we could predict as successfully as we could retrodict. The reason why we can trap ourselves (on the assumption that we are spectator-agents) and the reason why we can trap any spectator-agent in the above way is precisely that there is some sense in which a spectator-agent does not initiate his actions: they happen to him, and they do so (we can assume) in ways that are transparent to explanation along causal and rational lines. This allows observers to predict his future actions, and it also allows them to devise fixed points, such that when the prediction 'You will do Y' is fed into the agent's cognitive system it has the effect, when combined with other causal and rational grounds for action, of generating just the action Y. A spectator-agent trapped in this way cannot exploit the epistemological loophole; he cannot operate with an epistemic asymmetry between past and future, and so he cannot construct for himself - i.e. entitle himself to a conception of - anisotropic time.

This argument would be vitiated (because it would prove too much) if it were possible to construct such fixed points for free agents. But it is not. A free agent has the power to break out of

any particular set of predictions concerning his future actions, so that the epistemological loophole will always be available to supply such an agent with epistemic, and thence temporal, asymmetry. The hypothesized spectator-agent, on the other hand, cannot guarantee such an asymmetry for himself. He cannot rule out the possibility that he might be manipulated - or manipulate himself - in such a way that future and past would be epistemically symmetrical for him. He cannot prevent the mutual assimilation of past and future, and so cannot preserve a necessary directedness in the way he thinks about time; hence he cannot preserve directedness at all. Assuming that epistemic asymmetry is a necessary feature of any agency competent to direct time (an assumption which seems to me overwhelmingly plausible: what else could the difference between past and future consist in?), and given also that compatibilist agency cannot secure necessary epistemic asymmetry in the way it conceives of agenthood, it follows that compatibilism is not competent to direct time, and so not entitled to help itself to time at all. That is a reductio of the compatibilist's case.

I claim that it is overwhelmingly plausible that necessary epistemic asymmetry is at least partly constitutive of the direction of time, and the above argument rests on this claim. Can we say anything in its defence? I think we can. Suppose we consider a spectator-agent who is not, to his knowledge, being manipulated in the described way, so that he takes himself to have knowledge of the past but not the future. Although such a spectator-agent has only at best a de facto and not a necessary epistemic asymmetry, one might wonder whether the de facto asymmetry is not adequate to enable him

in his position to direct time. Unfortunately, however, this agent is vulnerable to the following devastating scepticism: he cannot rule out the possibility that the cognitive faculty which he takes to be the faculty of memory is in fact one of something like precognition, and that what he takes to be ignorance of the future is in point of fact ignorance of the past, resulting from perpetual and blanket amnesia.

For suppose he has the impression of taking tea ten minutes away from now in one direction. (I make it precise for convenience.) How could he decide whether his impression is a memory-impression or a quasi-precognition? It is no good his 'waiting' ten minutes and seeing whether he then takes tea or not. Whichever course of action he follows ten minutes from now, nothing settles the issue between the rival hypotheses stated above. Nor can he decide to falsify the hypothesis that the impression is a quasi-precognition by not taking tea 'in ten minutes time'. His ability to make such a decision is precisely what has been set aside in establishing him as a spectator-agent in the first place. Recall that we are operating at the level at which he can at best observe himself 'deciding' (not) to take tea: he is a spectator-agent, not an agent simpliciter; he cannot decide simpliciter (not) to take tea. That we are entitled to operate at this level, and characterize it in something like this incompatibilist way, is a fundamental parameter of the discussion. I have defended it above in my discussion of deliberation, and it is a thesis which is defended in many incompatibilist writings on the subject, and one whose truth I am taking for granted here. Given that for any level at which the spectator-agent can be correctly

described as deciding or attempting (not) to take tea, there is always a deeper level at which the correct characterization of his action is that he is the spectator of his action and not the performer of it (because not the chooser of the performance) - given this, it follows that, at that deeper level, whichever action our envisaged spectator-agent finds himself performing ten minutes away in one direction from now, that action is compatible with either of the rival hypotheses: whether or not he takes tea at $t \pm 10$ minutes, the impression enjoyed at t - if veridical - could be either a memory-impression (and he is ignorant of the future), or a quasi-precognition (and he is amnesic concerning ten minutes ago).

Ten minutes away from now could, for the spectator-agent, be ten minutes either in the future or in the past; he has no way of telling which. The free agent, on the other hand, can decide not to take tea ten minutes from now in just one direction. If he has a veridical impression of taking tea ten minutes from now, that impression must have reference to the other direction. The possibility of making decisions in one direction in general rules out the availability to the free agent of impressions relating to that direction; so that all impressions, whether or not veridical, must aim to have reference to the other direction. The free agent can escape the scepticism to which the spectator-agent is liable. This argument does not have to do with epistemological considerations relating to the settling of questions of veridicality. All impressions - mnemonic or precognitive - aim to be veridical. That is their function. The point is then that the free agent can say of any impression he enjoys at an instant: if this

impression is veridical, it must be a memory impression. The spectator agent - and hence the agent as the compatibilist conceives him - cannot lodge this claim. As far as he is concerned, the impression, veridical or not, might just as well be a quasi-precognition.

This scepticism is ruinous to the spectator-agent's chances of constructing an epistemic asymmetry between past and future.²⁵ In an extremely radical case of failure, such an agent might be equipped both with memory impressions, and with impressions (supplied by the manipulator) precisely matching the content of his future impressions of experience and action (inverse memory-impressions, or quasi-precognitions). Such an agent would have no means of giving sense to the idea that his impressions of events in one direction were impressions of the past, and his impressions of events in the other direction impressions of the future. For what could this difference consist in? He would enjoy complete epistemic asymmetry, and so would not be able to direct time. If he tried to conceive of himself as moving in time, he could not suppose that he was moving into the 'future' as opposed to the 'past', and so his attempted conception would be abortive.

As we have seen, the same problem confronts a spectator-agent in the less radical position of only having impressions relating to events in one direction. This agent cannot construct an asymmetry, because he has no means of distinguishing between on the one hand the hypothesis that his impressions are memories and that his ignorance in the other direction is ignorance of the future; and on the other hand the hypothesis that his impressions are

quasi-precognitions supplied by a manipulator, and his ignorance in the other direction is ignorance of the past, produced by a blanket amnesia caused by a manipulator. The spectator-agent cannot distinguish between the rival hypotheses because he cannot conceive of what the distinction would consist in. So the supposition that just one of these hypotheses is true is empty.

The free agent, on the other hand, enjoys a necessary epistemic asymmetry guaranteed by (at least) the necessary status of the epistemological loophole in his actions. He cannot know about his own future, because fixed points cannot be constructed for his future actions. His ignorance in one direction is a necessary ignorance; whereas the spectator-agent's ignorance - in the less radical case where he had it - was purely accidental. Any epistemic asymmetry for him could at best be apparent: nothing in his position would be essentially changed were he to be supplied with veridical impressions in both directions. The past and future are two realms to which the spectator-agent has, in principle, epistemic access; if in point of fact he lacks knowledge in one of these areas, well, that is just an accident. I conclude that only incompatibilist freedom is competent to direct time.

It might be objected here that the spectator-agent can simply perceive temporal precedence; that temporal precedence is part of the content of any representational experience. For surely, it might be said, all we need to do is equip him with a specious present, that is to say a present which extends somewhat to either side of the instantaneous, so as to incorporate the duration of short-lived processes.²⁶ It is certainly plausible to suppose that temporal

precedence is part of what is perceived in a specious present: the subject takes himself to enjoy one temporally extended if short-lived experience, rather than a series of snap-shot instantaneous experiences. Of course the above argument depends on his experience being instantaneous; if we could just award him experience of the specious present, we would thereby accord him experience which would straightforwardly incorporate temporal precedence as a bottom-line feature of the representationality of his experiences.

But it is very doubtful, to say the least, whether the spectator-agent can simply help himself to the perception of temporal precedence in this way. The ability to enjoy perception of temporal precedence in a specious present does not come apart from abilities to remember the immediate past and to anticipate the immediate future.²⁷ That is because the specious present does not have sharp parameters. It shades off into the past and future; and hence the ability, for example, to perceive an earlier segment of a speciously present process as earlier shades off into the ability to remember events which are so far anterior to the present moment as to be quite definitely in the past. The ability to perceive temporal precedence in a specious present therefore requires, at least, the ability to exercise the concept of memory, and that in turn requires the ability to distinguish in thought between memory and precognition. And we have seen that the spectator-agent does not have this ability.

Putting the point directly, instead of indirectly in terms of the epistemic asymmetry between memory of the past and anticipation

of the future, we might say that in order to perceive an event (say, hear a tune) in a specious present, it is necessary to distinguish between earlier and later phases of the event; and that ability requires one to suppose that one could, as free agent, interpose after hearing an earlier segment of the tune to prevent the occurrence of the later phase. To relinquish the freedom to intervene in this way would, as we have seen, be to relinquish the ability to distinguish in thought between earlier and later, and hence to destroy not only a grasp of the specious present, but any grasp of time at all (since the idea of an instantaneous present is no more available independently of the idea of past and future than the idea of a specious present). The supposition that one might extract the notion of an objective time-order from the bare notion of the order of perceptions is what is at fault in the above argument. Something else needs to be added.

Kant argues in the Second Analogy that causation in the objects of perception supplies the needed extra. The point is supposed to be that a perception of A followed by a perception of B can count as a succession of perceptions of an objective change only if the order of events is fixed in the direction from A to B. But it is important to see here that the argument proceeds by a non sequitur, as Strawson rightly accuses it of doing.²⁸ Strawson objects that necessity in the order of perceptions does not imply necessity in the order of objects. For given that A follows B, the order of perceptions is fixed, but nothing follows as to the connection - necessary or otherwise - between A and B. One might suppose that unless the order of events is necessary, the order of

perceptions cannot enjoy the necessary order they need to have if they are to supply the subject with a conception of objective change. But this is incorrect. It does not follow from the fact that the perception of B could not have occurred before the perception of A that B could not have occurred before A. Strawson's point is that a contingent order of events (say, A before B) might fix a necessary order of perceptions (A before B); the order of perceptions is necessary given the (contingent) order of events. So the subject can think of the order of his perceptions as necessary (and hence objectively valid), without thinking of the order of events as necessary; to follow the Kantian line one must be prepared to suppose absurdly that whenever one perceives any A followed by any B, A and B are causally connected. Kant indeed embraced the conclusion of universal causal interaction, or community, but there is little temptation to follow him in this.

The general form of Kant's argument - that fixture in the order of perceptions requires fixture in the order of events - is thus unsound. What is also incorrect is to suppose that causality can fix the order of events uniquely; I have argued that it cannot do so. (Kant supposed that it could do so, but his 'community' hypothesis - which includes reciprocity in the causal connections between events - crucially undermines the supposition.) What is needed, rather, is the possibility of free intervention in the course of events. A subject can conceive of the succession of his perceptions as revealing to him a succession of events only if he can conceive of the events as anisotropically directed. And he can achieve this conception only if he can suppose himself (or anyone)

in principle capable of intervening to prevent the occurrence of those events certified as occurring later by the superimposition of the anisotropic time-direction. The objective time-order is secured by the impossibility of altering those events in the observed process that are now past, and the possibility of affecting those events which will occur beyond what has been so far observed. Only incompatibilist free agency can earn this distinction between past and future, and thereby entitle itself to a conception of A-series time, and hence time itself.

3.4 Freedom and Causality: the transcendental nature of the
A-series.

Freedom underwrites the direction of time, and the direction of time supplies causality with its direction. We can now draw a deep connection in another place between freedom and causality. As I have suggested, the primary manifestation of causation in rebus lies in the exercise of causal dispositions by things of certain sorts possessing those dispositions. To have a causal disposition is to have the power to behave in certain ways in certain circumstances. Now clearly these circumstances cannot be restricted to actually obtaining circumstances: counterfactual circumstances also come within the purview of causal dispositions. There is no room to insert an ontological wedge between the actual and the counterfactual here. To say that something is an egg is to say not only that it will break if dropped in normal circumstances, but also that were it to be dropped, or had it been dropped, in normal circumstances, then it would, or would have, broken. This anti-empiricist idea of things of certain sorts with certain causal dispositions, which reach out from the actual into the counterfactual, is fundamental to our metaphysical picture of the world.

What supports the notion of a causal disposition? I suggest that nothing short of full incompatibilist freedom can do so, because only such freedom can supply the idea (integral to the notion of a causal disposition) of ways things might have been. This

is because it supplies the idea that, given any time designated as now, the past relative to that time is fixed, and the future open. Since we can accommodate the existence of times which are in point of fact not now, but have been or will be present, we can make room for the idea that there have been open futures (now closed), and will be open futures (not yet open). In so doing we allow for the possibility that there are ways things might have gone other than the way they in fact went, and ways things might in future go, other than the way they will in fact go.

Mackie expresses something close to this point²⁹ when he claims that too much causation destroys itself. If everything were a thing of a certain sort with certain causal dispositions there would be no room for freedom, and hence no room for a distinction between the (necessarily fixed) past and the (necessarily open) future. Causes and effects would be jointly fixed, and there would in principle be no warrant for regarding events in rebus as exhibiting causal asymmetry. The designations of 'cause' and 'effect' would be interchangeable without loss of explanatory power. What secures the asymmetry of the causal relation is the possibility of causes being fixed before effects are fixed. And what secures this possibility is the possibility of free intervention in causal processes in rebus so as to bring about effects by means of implementing causes, and preventing 'effects' from occurring by preventing processes from running forward after the 'cause' has occurred. It follows that the agent with power freely to intervene in causal processes cannot himself be a thing of a certain sort with certain causal dispositions. The intervening self must be transcendental with

respect to the causal world-order.

We should be careful here concerning the exact status of the unfixed effect. In any case of causation, it is natural and, I shall take it, correct to follow Mackie in saying that, given the circumstances, the cause is necessary and sufficient for the effect. The implications here between cause and effect are not material implications, but include both the truth of the antecedents and the truth of the related counterfactuals. A caused B iff, given the circumstances, A occurred iff B occurred; A and B occurred; and A would have occurred if B had been going to occur, and B would have occurred if A had occurred. It follows that given the circumstances, or what Mackie calls the 'causal field',³⁰ there is, as we ordinarily conceive causal relations, no asymmetry between cause and effect. This does not mean, however, that there is no asymmetry between cause and effect period, because the given circumstances just include the fact that no interference in the process occurred; the asymmetry is then guaranteed by the fact that it was possible to intervene in the process after A occurred to prevent the occurrence of B. There was a time when the circumstances were not entirely given.

We are now in a position to say exactly how the direction of time constitutes the direction of causation. Causation is an asymmetric relation: if X causes Y, Y does not cause X. But the asymmetry, and hence the causation, between related objects or events has to be earned: it is not simply given in an ordering of objects and events. What needs to be secured is the possibility, for at least some such objects and events, of regarding some of those

items as fixed before others of them are fixed. That possibility, I have argued, cannot be supplied by the mere idea of a thing of a certain sort exercising certain causal dispositions. It can only be supplied by adding the idea of things not of a certain sort not exercising causal dispositions, but nevertheless possessing the power to intervene in and manipulate objects and events in the world. Only such things can supply the openness of the future and the fixity of the past. Causality depends for its direction upon the existence of free agency, as well as depending on it for the prevention of the self-destruction syndrome. More articulately, causality depends for its existence and direction on the availability of A-series time, and A-series time is in its turn only provided for by free agency. Although the direction of causation and the direction of time necessarily coincide, metaphysical priority attaches to the direction of time. Causal priority has to be constituted out of temporal priority.

A-series time is transcendental in two connected respects. The first of these we have already noted. The temporal point of view which constitutes A-series time does not enter into the specification of the content of temporally indexed utterances or of propositional attitudes needed to account for behaviour. This transcendence is quite general to the indexical. If I find a bottle washed up on a beach containing the message 'I am here now', the message is in itself contentless. Of course information can be deduced from the existence of the message and its likely provenance, but in itself it says nothing: nothing further can be deduced from the message than could be deduced from a similarly encountered piece

of paper marked with illegible scrawls. Likewise, if someone goes to vote because he realizes that election day (say, June 11th) is today, while the content of his realization is that June 11th is the day he is at, and while the content of that realization cannot be specified tenselessly, tense does not occur ineliminably in the specification of the facts we need to cite in order to account for his voting behaviour. In particular, his going to vote on June 11th is accounted for by his desire to vote and his belief that June 11th is polling day. His belief, on June 11th, that June 11th is today has admittedly no tenseless surrogate, but is explanatorily otiose.

How should we specify what it is for June 11th to be today? There are several ways in which this might be done. Earlier, I gave one possible such specification: June 11th is contemporaneous with this mental state, where 'this' demonstrates any mental state the subject is enjoying on June 11th. Of course 'this' here just means 'the one I'm having today', so no reductive analysis is offered by the specification. The case is the same with another possible specification, which is available in the light of our above argument for the necessary role of freedom in constituting the A-series. What makes it the case that June 11th is today (at, say, the moment when it becomes today) is the fixity of events up to and including June 10th, and the openness of events from and including June 11th onwards. Now this specification of the todayness of June 11th is more essential than the specification in terms of contemporaneous mental states, for it appeals to the very thing which constitutes the todayness of June 11th, rather than a merely adventitious feature of it. (This mental state - the one I'm having today - might

have failed to exist; but today would not in that case have failed to exist.) But both specifications are equally tenseless. The informative content of any tensed sentence can and must be captured tenselessly.

The second respect in which the A-series is transcendental then emerges as follows: there is no explaining why a time *t* is now. There is, as we have said, a tenseless explanation of what it is for *t* to be now: that is supplied by the anisotropy of time, which is underpinned by freedom. For a time *t* to be now is for an agent to be free to affect events at and after *t* but not before. But such a tenseless specification is obviously available to constitute every moment as a now, and of course indeed does so: every moment is, in its time, now. But the tenseless explanation of what the nowness of a moment consists in does not discriminate between moments. Tenseless statements do tell us which moment is now, since only one such tenseless statement is true at any given moment. But there is no explaining why the particular tenseless statement which is, at any moment, true is true, rather than another such statement indexed to another moment. Why is June 23rd today (as I write this, it is) and not June 22nd? Well, June 23rd is today; that fact stands, but inexplicably. It transcends explanation because it transcends informative content; the informative content of any purported explanation could not reach as far as that fact, and so could not explain it. The fact that June 23rd is today is a transcendental fact.

This second respect in which A-series time is transcendental is related to the first respect. As we noted at the beginning of

this chapter, the A-series is continuously on the move: the present moment moves along the B-series like a spotlight illuminating one date after another. Of course, as has been frequently pointed out against Newton's notion of absolute time, it is paradoxical to say that time moves or flows. Time cannot flow because there is no rate at which it flows. (If there were it would have to be the rate of one unit of time per unit of time, which is nonsense.) But it is not paradoxical to say that the A-series is continuously on the move. That is just to say that the present moment is continuously shifting. But again this fact is transcendental, and is another aspect of the transcendental fact noted above. For just as there is no explanation as to why a time t is now, so there is no explanation of the movement of the present moment. The two phenomena go together. We could explain why a moment is the present moment (when it is) if, and only if, we could explain what it is for the present to move. For a moment is present precisely when that is where the moving now has got to.

The anisotropy of time which free will secures guarantees that each moment of time, when it is present, has a unique future and past such that there is a genuine distinction between the future and the past at that moment; it is a further consequence of the role of free will that only one moment at a time can be present - but just so far these are B-facts: freedom does not yield the A-fact that one moment is absolutely present, namely this one. One might try to specify what it is for the present moment to move in terms of the necessarily shifting domain of free action: once an action has been performed its performance is fixed. But again, this specification at

most amounts to the B-fact that at any instant I can affect what happens at and after that instant but not before; it does not yield the A-fact that just one of those instants is (absolutely) now. Nothing in the specification of free action itself secures for it a shifting (rather than a merely relativized) domain. Freedom secures anisotropy, but I do not see that as such it secures movement; it could not do so, because the nature of free action is tenselessly specifiable, whereas both the movement of the present moment and what it is for a moment to be (absolutely) now are transcendental facts: they have no tenseless surrogate.

But freedom to act in time is freedom to act in A-time, in the fullest sense. Since the freedom of the self is what constitutes the anisotropy of the A-series, the transcendental status of the self with respect to the causal order must in some way constitute the transcendental status of A-time in both of the two senses we have specified. How the transcendental status of the self enables and supports the transcendence of A-time in the second of the above senses (the inexplicability of why a moment is now when it is now) is to me entirely dark; but that it does so in some way looks plausible. I shall suggest in the next section how we might understand the way in which the transcendental self supports the transcendence of the A-series in the first of the above senses: the irrelevance of A-time to the project of accounting for behaviour.

3.5 Freedom and the Transcendental Self

I want now to connect the previous discussion with more general considerations relating to the self. It has recently been argued by Galen Strawson³¹ that genuine, incompatibilist freedom is not possible, at least not given that there can be full explanations of free action. Free action is at least rational action, and rational action is such that an explanation can be provided for the action by citing the reasons which moved the agent to perform it. If such explanation can be, as we often believe it is, complete, no room is left for any indeterminism to come between the reasons and the action, and so no room is left for a free agent to exercise choice. It is important to note that Strawson's argument does not rest on a causal theory of action, and so is free from the traditional terms of the dispute between libertarians and determinists. The point is that explanation of rational action in terms of reasons for action leaves no explanatory gap into which freedom might be inserted. It might be thought that freedom could be inserted earlier in the explanatory chain: perhaps the free agent can choose the reasons even if he cannot choose to act once the reasons are in place. But, as Strawson correctly observes, this manoeuvre is hopeless since it leads to a regress. If the choice of reasons is to be such, it must be rational; that is to say, the agent must choose his reasons for reasons, and so the problem just recurs. Furthermore some of the agent's reasons will not be so open

to choice: they will come in the form of beliefs and desires which are just brutally and inexorably given to the agent. But even where the agent chooses to act on one desire rather than another, say, the choice cannot be genuinely free: it must be made on the basis of other beliefs and desires in the agent's repertoire. And so on. The general shape of the argument is that to be free an action must be rational, to be rational it must be fully explicable, and it must be explicable in a way which leaves no room for freedom. So there is no freedom (Cf Kant B574).

Apart from the very last step, I believe this argument is sound. What it shows (extruding the last step) is that freedom transcends explanation, not that there is no such thing as freedom. Freedom does not supply a component in the explanatory nexus connecting reasons for action and action, but only a deterministic causal theory of action would be entitled to conclude from that fact that there was no such thing as freedom, and Strawson has rightly detached his argument from such a theory.

Now it might be thought that we could avoid this conclusion by allowing an element of inexplicability to enter into our accounting for free rational action. But I do not think this is a defensible manoeuvre. Consider an agent who freely and rationally chooses to perform A for reason R rather than B for reason R'. Now either the agent thinks of R as as good a reason to perform A as R' is to perform B, or he does not. If he does, and nevertheless goes on to perform A, say, then we certainly do have an element of inexplicability in our account of his action. For while we can explain why he performed A (namely for reason R), we cannot explain

why he performed A-for-reason-R rather than B-for-reason-R'. Ex hypothesis there is no further reason S which ranks R and R': on the present assumption the agent thinks that R and R' are equally good reasons for their respective courses of action. But rather than remaining in perpetual indecision like Buridan's ass, he goes on to perform A. There is therefore something inexplicable about his action. And for that very reason his performance of A cannot be regarded as a genuinely free and rational action. The inexplicability detracts from the action's rationality, and hence detracts from its freedom: only genuinely rational action can be free; or, at least, the more rational an action, the better candidate it is for the status of being free.³² As with Buridan's ass, the agent's choice of A rather than B would have to be regarded as random (and so not free, in the required sense). That does not make his action irrational, but it does make it not entirely rational. If, on the other hand, the agent does not rank R and R' equally, presumably preferring R, then he will, if he is rational, perform the action supported by R (namely A). In this case his action will be fully explicable, and explicable in a way which leaves no room for freedom. The conclusion is unavoidable that only genuinely rational action can be free, and genuinely rational action is fully explicable in a way that leaves no explanatory tasks for freedom to perform; those tasks are all picked up by the reasons for action. Freedom is transcendental with respect to explanation of action.

It is now time to connect up this discussion with that of the previous sections. We saw that freedom was needed to underpin the

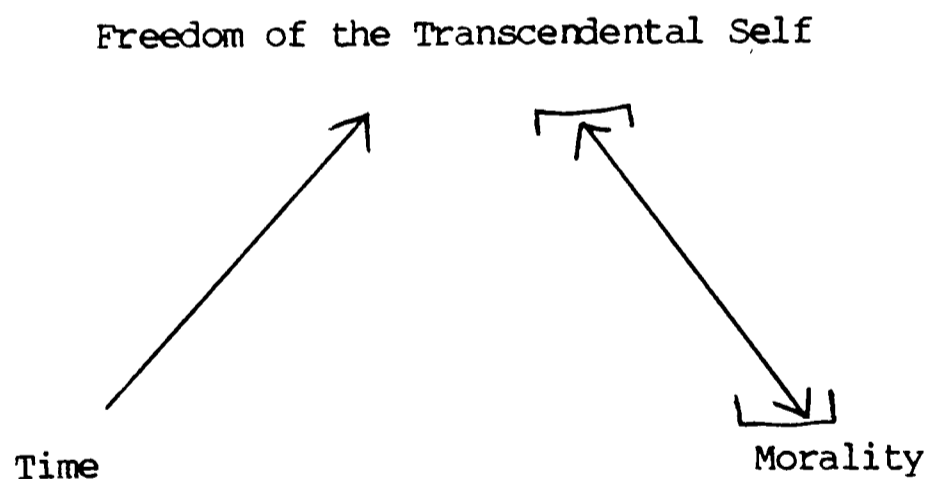
A-series, which is needed to constitute time. But we also saw that the A-series transcends language, and in particular transcends the explanatory project of accounting for action. Reasons for action can, like propositional content, be specified in purely B-terms. This suggests the following partially Kantian line of thought.

Freedom is the activity of the transcendental self in its capacity as agent (rather than as subject). It is not the activity of the empirical self, because that self is constituted by its network of beliefs, desires and other propositional attitudes; and, as we have seen, there is no place for freedom in this network. If there were a place for it, there would be a role for it in the explanation of action (since that is what this network is for), and we have agreed with Strawson that freedom can play no such explanatory part. The empirical self is also part of the causal order, and we saw in the last section that the self which constitutes time must stand outside that order if it is to be constituted at all. Freedom is transcendental with respect both to the explanation of action and to causal explanation. But there must be freedom, because there is time; that leaves only the transcendental self for freedom to attach itself to - the self which stands behind the empirical constituents of mind, the self which Hume couldn't find because he wasn't prepared to look beyond those constituents. We have drawn, therefore, an intimate connection between time and the transcendental self. The transcendental self is what makes time (and causation) possible. This suggests the Kantian doctrine of time as the form of inner sense. But the position I have reached differs from Kant's in at least two crucial respects:

firstly, Kant's transcendentalism incorporates, as mine emphatically does not, the claim that transcendental selves stand outside the temporal, as well as the empirical, order; secondly, Kant does not connect up the triad of freedom, time and the transcendental self in the intimate way I have tried to do. Rather, he relies on rationality - and in particular the rationality which is expressed in moral duty - to guarantee the freedom of the transcendental self, leaving the special connection between time and the transcendental self hanging as a loose end:

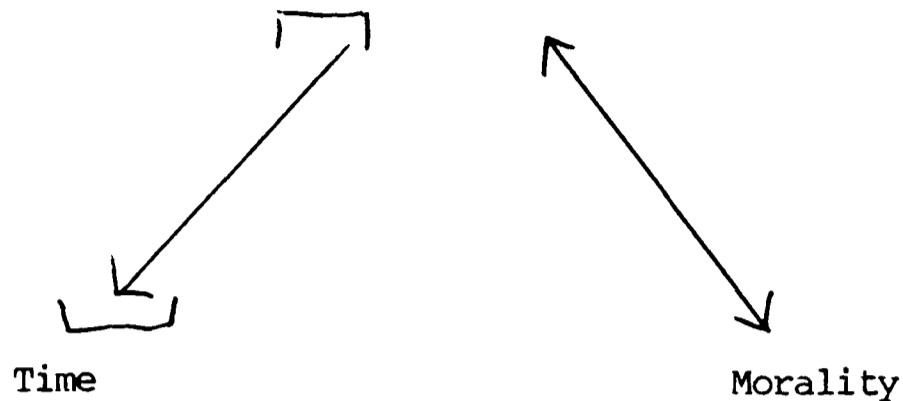
For had not the moral law already been previously distinctly thought in our reason, we would never have been justified in assuming anything like freedom, even though it is not contradictory. (Crit. Pract. Reas. Ak.5.)

The picture I am offering looks like this:



(I shall later argue for the mutual implication between time and the transcendental self, so that the complete picture, including the unargued connections with morality, would link all three elements in a relations of mutual implication.) I take it that Kant's picture is this:

Freedom of the Transcendental Self



Kant's idea is that freedom, along with the existence of God and the immortality of the soul, is, as he puts it, a postulate of pure practical reason. By this he means that it is presupposed by the moral law (op.cit., Ak.132). It is not an object of theoretical reasoning (or, as we might say, metaphysical reasoning) because it is not a possible object of intuition. Metaphysical reasoning is, in Kant's view, restricted to the a priori conditions of empirical intuition. Such reasoning does, indeed, leave a place for the conceivability of freedom.³³ But it cannot prove the actuality of freedom. Only the fact of morality can do that, because only this fact can assure us that there must be a component in the moral self which is independent of the causally conditioned empirical order:

For nothing more has been accomplished by practical reason than to show that those concepts are real and actually have (possible) objects, but no intuitions of them are thereby given (and indeed none can be demanded), and thus no synthetic proposition is made possible by conceding their reality. Consequently, this disclosure does not in the least help us in a speculative respect....Now through an apodictic practical law, they, as necessary conditions of the possibility of that which this law requires to be made an object, acquire objective reality. (Crit.Pract.Reas. Ak.134-5.)

We can agree with Kant that metaphysical reasoning cannot

deduce the existence of a free self such as is capable of being the object of a possible intuition. In that sense, we have no knowledge of the nature of the free self, but only knowledge of the fact of freedom (Ak. 133, 136). There is nothing other than this fact to find out about (Ak. 134). As it stands, this point is just truistic: if the domain of knowledge is restricted - as it is by Kant - to what can be an object of empirical or a priori intuition (where a priori intuition relates to possible objects of empirical intuition), then it is clear that the free self cannot be an object of knowledge, simply because the free self is outside the empirical order. So as Kant sets up the terms, the existence of the free self cannot be deduced by speculative reasoning.

But it is not mandatory to operate within Kant's restriction; relaxing it, we should reject the suggestion that the existence of the free self cannot be deduced by metaphysical reasoning. A metaphysical deduction which arrived at the free self would be an exercise in pure theoretical reason rather than pure practical reason; it would not seek to base the deduction on the existence of specifically moral agency, but rather on the existence of agency as such. Such a deduction would have the power to yield us an object - the free self - which would not be a possible object of intuition, because required to lie outside the empirical order. Theoretical reason would therefore be able to do more than take objects delivered by pure practical reason and 'think those objects by means of the categories' (Ak.136); it would actually be able to deduce their existence, and hence short-circuit the traditional and much-vexed Argument from Morality to the freedom of the will. That

Argument suffers from two drawbacks. Firstly, it is simply unclear whether morality does indeed presuppose freedom (Aristotle seems not to have required it). But secondly, even if there is a component of moral thought which presupposes freedom, it is unclear whether or not that component is indispensable. Perhaps something like morality could survive in the absence of a belief in freedom.³⁴ On the other hand, perhaps morality does presuppose freedom, so that they have to be preserved or jettisoned together. And perhaps too morality and agency are so closely connected that nothing recognizable as rational agency would survive the abandoning of a moral system. These are speculations which I cannot pursue here. (I incline to think that both of these latter suppositions are true.) Nevertheless, there is some argumentative value in attempting to derive freedom from the assumption of rational agency alone, without bringing morality into the question. The argument which I have offered has the form of such a deduction, and so is - if successful - a metaphysical deduction of the existence of the free self.

But Kant's point that the free self cannot be an object of empirical knowledge is not challenged by the considerations I have marshalled; indeed it is supported. The free self cannot have empirical characteristics, because things with empirical characteristics (things of certain sorts) behave predictably in certain sorts of ways in certain circumstances: they engage with the explanatory and the causal order. But the free self cannot do that without relinquishing its freedom. It is therefore transcendental with respect to the empirical order: in Kant's terminology (though without his implication of atemporality), it is a thing-in-itself.

In the above schematic representation of the argument I have been offering, time and the freedom of the transcendental self were linked by an implication from the former to the latter. I have not explicitly argued for the converse implication, but it is obvious that it must hold. Freedom to act is freedom to act in time; and it must be freedom to act in just one temporal direction, because freedom to act in both temporal directions would undermine the possibility of acting in one direction. The agent free to act in both directions could, for example, act in one direction so as to destroy some of the beliefs and desires on the basis of which he acts when he acts in the other direction; which is a contradiction. (The beliefs would perhaps be easier for him to alter: that would merely require him to change those features of the world onto which the beliefs in question are directed.) At the extreme, such an agent could commit suicide in both directions; which is again a contradiction.³⁵

The epistemic asymmetry which, as I have suggested, is a necessary concomitant of time's asymmetry, can be viewed as a corollary of this. If an agent can act in one, but only one, temporal direction, it follows that the events lying along the other, but only the other, direction are potentially cognitively available to him. Conversely, an agent who could act in both forward and reverse directions would not have a past and a future: rather, he would have two futures. But then his cognitive attitudes to his two futures would be epistemically symmetrical. He could not know about either future, because that would debar him from bringing it about. Contrariwise, if an agent does know about events lying along

one temporal direction, say by the exercise of memory, he cannot conceive of himself as having the power to affect events in that direction. If he has cognitive access to any events in that direction, he must conceive of himself as potentially having cognitive access to all of them, because they are epistemically on a par. So he must regard them as fixed, that is as not now negotiable. It seems then that we must find a mutual implication between freedom and temporal epistemic asymmetry; more generally, between time and the existence of the ^{free} transcendental agent. That is the principal conclusion of this chapter.

This argument rules out the possibility of backwards causation, in the sense of causation which runs against the temporal direction. If there could be backwards causation, it would be possible to manipulate such causal processes to bring about past effects. But the temporal asymmetry between past and future yields an epistemic asymmetry: the past, being knowable about, is fixed. It is not possible to change it. So a causal process initiated now cannot affect the past. This argument trades crucially on the epistemic parity between those past events which are known about and those which are not. There can be no dispute that such of the past as is known about is not now alterable; the argument then insists on an epistemic uniformity between the known past and the unknown past: anything in the past which is not known is nevertheless knowable (in principle). This in turn requires the topic-neutrality of the ontology of the past: past objects and events existed in exactly the same sense. This is an objectivity requirement: it requires past objects and events to inhabit the same objective world. The

objectivity requirement is met in virtue of two facts. The first of these is the (at least inchoate) objectivity of the world as experienced in present-tense representational experience, and secondly the recursive nature of memory: for to remember is to remember having seen (or heard, etc.).³⁶ The exercise of memory depends upon the earlier exercise of perceptual or sensory faculties, so that a memory in effect calls up an earlier non-mnemic experience. In the case where that experience was representational (as when I remember having seen, heard etc.), the objectivity of the previously experienced items simply carries over to the memory. If we have any access to the past through memory, we have access to an inchoately objective past.

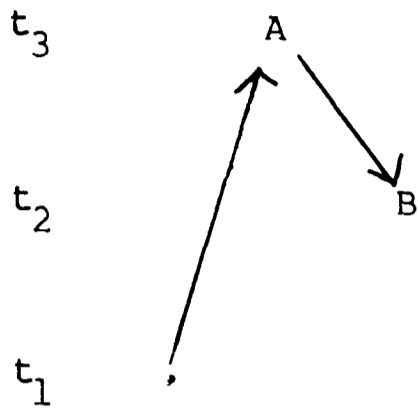
Anti-realism about the past goes wrong by falsifying the content of memory experience, since it makes out that the connection between a now present memory and the remembered past is at best evidential. In the same way, anti-realism about present-tense experience falsifies its content, by suggesting that such experience gives us something less than a world of objects (gives us mere evidence for it). If we insist that nothing can amount to a representational experience which does not meet the requirement of inchoate objectivity, we can equally insist (on the basis of the recursive character of memory) that a memory which exploits an earlier representational experience is not available independently of the (inchoate) objectivity of the remembered items. The inchoate objectivity of those items carries with it their topic-neutrality, and it carries with it too the idea of the world, housing those objects, all of which - experienced and unexperienced - enjoy

ontological parity. And if ontological parity, then epistemic parity. Hence backwards causation is not possible, for that would require some of the past not to be fixed, and so not now knowable about, while other areas of it were fixed; which contradicts the epistemic parity which our argument shows must attach to past objects.³⁷

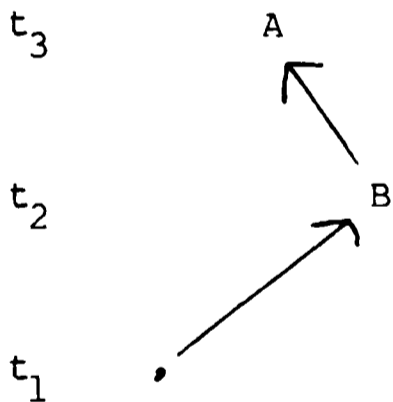
Of course we cannot, without begging the question, appeal to epistemic asymmetry on its own as a consideration in favour of the epistemic parity of past objects, and so as a consideration against the possibility of backwards causation. Plainly, if we were forced by the production of overwhelming evidence to accept cases of backwards causation, that would eo ipso undermine our belief in epistemic asymmetry. Epistemic asymmetry and the bar on backwards causation mutually support one another, so that we cannot use one to gain leverage on the other.³⁸ But I have not tried to do so. I have argued that epistemic asymmetry is a corollary of freedom, which is itself a condition of temporality. So the rejection of, in the first instance, anti-realism about the past, and in the second instance, the possibility of backwards causation, is firmly anchored in the nature of temporality. Epistemic asymmetry is a link in the chain; but it is not itself performing a grounding function. The epistemic asymmetry arises from the fixture of the past and the openness of the future, as constituted by the freedom of the agent.

It might be thought that a more sophisticated form of backwards causation could be argued for in the following way: while it is conceded that it is not possible to affect the past simpliciter, perhaps it is possible to affect the past in the future, by

affecting the more remote future and exploiting a backward causal process. The picture would look like this:



At t_1 it is brought about that A will occur at t_3 , and thereby that B will occur at t_2 , via a backwards causal chain from A to B. The initial difficulty with this supposition lies in supplying a reason why we should ever accept this account of the matter, rather than a 'forward' version looking like this:



But the deeper difficulty is that the proposal still does not escape the main objection to backward causation. Although it is true at t_1 that B is not fixed while A remains unfixed, it becomes true at t_2 . There can be no objection in principle to implementing the following procedure (the bilking manoeuvre): letting things run on to t_2 , observing the occurrence of B (say), and then preventing the

occurrence of A, if necessary by destroying a sufficiently large segment of the universe.

The direction of time and the direction of causality are, therefore, the same not only in general but without exception throughout the causal order. But the former is, as I have argued, more fundamental than the latter: it fixes the latter. It might be objected here that the direction of time cannot fix the direction of all causal relations, because it cannot fix simultaneous causal relations (which are temporally undirected). Consider Kant's case of a leaden ball making a hollow in a cushion. What makes it the case that the causal direction runs from the ball to the hollow and not vice versa? Surely the direction of time cannot help us here?

It is quite true that the direction of time cannot help us directly. But it can do so indirectly. Recall that the direction of time is required to select one set of causal dispositions from a pair of such sets possessed by each object in advance of the fixing of the time-direction. In the case of the cushion, one set contains the disposition, among others, to receive the imprints of hard objects, and the other set contains the disposition to propel such objects out of indentations in which they sit, in so doing recovering an integral shape. Now the time-direction settles which of these sets is activated, and thereby it settles which direction causation runs in in Kant's example. The essential point is the anti-Humean one that causation is not constructed out of, or built upon, mere regularity, but presupposes it. That is because causation is primarily the exercise of causal dispositions, and such dispositions contain the seeds of regularity. The causal relation

between the ball and the hollowed cushion is not, causally speaking, a one-off event, but rather a particular activation of the causal dispositions of leaden balls and stuffed cushions. These dispositions extend factually and counterfactually to other encounters between things of these sorts, and they also extend - more importantly for the present point - to the possible antecedents and developments of the present situation. Given a time-direction, the cushion is determined as having been deformed by the ball, and not as about to recover its integral shape by propelling the ball out of the hollow. I conclude, then, that the time-direction can supply simultaneous causation with its direction.

3.6 The Phenomenology of Freedom.

What is the connection between believing that you are free and being free? Consider first the question whether being free is a necessary condition of believing that you are free. If you believe that you are free, you must conceive of yourself as existing in time, since, as we have seen, the notion of freedom is not available independently of the notion of the time-series. You would also have to have certain basic manipulative and motor abilities: you would have to conceive of yourself as performing actions directed onto the world. You would have to do things which made a genuine difference to the spatial or quasi-spatial arrangements of objects, and conceive of yourself as having the ability to choose exactly how you manipulate those objects. Finally, to earn the belief that you are free, you would (as with other beliefs) have to be a sufficiently rational agent.

When the picture is fleshed out in this way it becomes very attractive to say that belief in freedom is sufficient for freedom. For what extra element could be required to secure it? Surely a rational agent who has the cognitive resources to entertain a genuine belief to the effect that he is free, and who does believe that at least some of his rational actions are free, has all that is needful to constitute him as a genuinely free agent. It is no objection to say that his actions may nevertheless be explanatorily determined. That does not militate against his freedom: we have already conceded that freedom cannot be fitted into the explanatory

picture. I cannot see, then, that anything more is required for the freedom of a rational agent than the genuine belief that he is free in at least some of his rational actions.

Galen Strawson argues, in the work already referred to, that our experience of freedom is illusory. In so doing he relies on a general principle to the effect that regardless of what appears to be the case, it always remains a possibility that reality is different. This allows him to conclude from the fact that freedom cannot be fitted into the empirical world that there is no such thing as freedom; even if it is a necessary illusion (one which unavoidably accrues to all rational agents), it remains an illusion. But the general principle cannot be allowed to stand. There are well-known cases where it breaks down, for example where an illusion is self-fulfilling: if one thinks that one is under an illusion (that one is making a mistake), then it follows that one must indeed be under an illusion (that one must be making a mistake); either in the place where one thinks one is going wrong, or, if not there, then in the supposition that one is going wrong in that place. More relevantly to our purposes, the general principle conflicts with the following highly plausible claim: if one cannot supply any account of what it would be to lack an illusion, the hypothesis that one is under the illusion loses content. In the case of many optical illusions we can supply such an account: we know what it would be for Frazer's Spiral to lack its illusory effect - it would be for perceivers to see it as a series of concentric circles (which is what it really is) rather than a continuous spiral (which is how it looks to normal perceivers). But what would it be to lack the

so-called illusion of freedom? I have argued that it would involve nothing less than losing our belief that the world is in time; that is not a belief we can imagine giving up. At best we could say that freedom is an illusion common to all rational agents; but to say that would in effect be to say that it is a condition on rationality in an agent that he have this illusion; and that is to say that freedom is just an illusion (if it weren't an illusion, it would be possible rationally to believe that it is not); and that is to say that time is an illusion (by a plausible contraposition); which is unacceptable. Equally, since freedom attaches to the transcendental self, if we are not prepared to regard time as an illusion, it loses content to say that the existence of the transcendental self is an illusion.

It might be conceded that, in general, belief in freedom is sufficient for freedom, but argued that in a particular case belief in freedom is not sufficient for freedom, because, an agent can be subject to constraint. An agent may believe that he can do either of A or B, but it might be the case that were he to attempt to perform B, his effort would be frustrated. If the frustration were external to his act of will, we would in such a case surely restrict the scope of action to the willing: he would not be free to do either of A or B, but he would be free to will to do either of A or B. But suppose now that he is subject to some kind of psychical constraint, so that although he believes he can will to do B, were he in fact to attempt to will, he would fail. In that case would we not say that while he believed he was free, he was not in fact free?³⁹ This sort of case is somewhat hard to assess. One might take the line that

such cases must necessarily be exceptional, thus seeking to preserve a general implication from belief in freedom to freedom, but it is difficult to see how to justify such a restriction. A more promising approach would tackle the locution 'were he to attempt to will...'. If there is such a thing as attempting to will, then we can simply re-apply our above argument for the case of external constraint to the case of internal constraint. The latter sort of constraint would in fact emerge as a species of the former. In other words, someone under psychical constraint would still be free to attempt to will, although not free to will. So belief in freedom would in both cases of constraint (external and internal) be sufficient for freedom, although the subject would not necessarily enjoy freedom in the place where he thinks he does.

In the case of psychical constraint, the free actions in question would be the attempt to will to do A and the attempt to will to do B: the agent would still be able to make either of these attempts. But clearly the threat of regress looms. It is open to the opponent of the subjectivist thesis to push the constraint in yet further, and argue that there could be cases where were the agent to attempt to will to do B he would fail. The subjectivist could then counter with a pair of alternative actions at the same level, either of which would be available to the agent: namely the attempt to attempt to will to do A (B). So if his opponent takes the dialectic into the regress, the subjectivist wins at every stage. But it will be said that there must be something wrong with the dialectic here, because the regress quickly takes us to mental states the intelligibility of which is extremely doubtful, to say the least.

Plausibly, there is willing to do something (or trying to do it), under normal conditions, and there may be such a thing as attempting to attempt to do something, under conditions of psychical constraint; and of course there is doing it (willing successfully); but that is all. If the agent wills (tries) successfully, it is usual to identify the action with the whole performance; if the attempt is a failure, we identify the action with the willing (the trying, the attempt).

The issue then comes to this: is it possible for an agent to believe that he can will (try) to do B and be wrong? Or rather, avoiding the regressive overtones one might detect in the phrase 'can will': is belief that there is genuine choice confronting one sufficient for there being genuine choice? Well, what more could be required? It is no good to reply to this question with the formulation: were he to will to do B he would succeed. For we must ask: succeed in what? He does not have to succeed in doing B; we have agreed that free choice does not require the absence of constraint. (If I deliberate over a course of action, I am not absolved from moral assessment, if it turns out that, owing to some constraint, I could not in fact have performed that action.) But it is trivially true that were he to will to do B, he would succeed in willing to do B. So the formula does not supply the extra purportedly needed to turn belief that one is free into freedom. And, as we have already observed, no version of the formula 'were he to...he would succeed' which appeals to higher-order attempts or willings can supply the missing extra, because at each stage alternative actions are available to the subjectivist without appeal

to the next stage of the regress. For example, suppose his opponent tries the formula: were the agent to attempt to will to do B, he would succeed in willing to do B. The subjectivist can simply reply that this success is not necessary for genuine choice. At the level in question, the agent has a choice between attempting to will to do A, and attempting to will to do B. If the opponent moves into the regress, taking on board its increasingly dubious ontology of mental states, the subjectivist has a riposte at every stage of it.

The subjectivist's opponent might try the following as a last-ditch manoeuvre. Perhaps an agent can believe that he is free, but be subject to so much constraint, that he can make no essays at all of any kind in the relevant field. The assertion here that he can make no essays of any relevant kind is not to be cashed as any non-vacuously true assertion of the form 'were he to attempt to..., he would fail'. The position is exactly that he can make no attempts. The subjectivist, it might be said, won the above skirmish simply because his opponent was presented as agreeing that whatever our agent may fail to do, he can always attempt to do it. That concession allows freedom to get into every act or failure to act at some level or other. But perhaps the concession should not have been made. Perhaps we should say that an agent can be in the position of just being radically unable to attempt anything (while believing that he can do so). That would still leave statements of the form 'were he to attempt to...he would fail' true, but true not in virtue of the non-vacuous truth of the subjunctive statement as a whole (as the subjectivist in effect assumed above); rather, vacuously true in virtue of the impossibility of the antecedent. For the fact is that

he cannot make any attempts.

This is an interesting line, but I do not think it works against the subjectivist. The difficulty with it is simply that it is in danger of leaching the content out of the hypothesis - that the agent believes that he can act freely. What would it be to have that belief, if one could not make any attempts with respect to the field of action in question? How could one be credited with the belief that he could, say, rise from his chair (or try to rise from it), if he could not attempt to (or attempt to attempt to) rise from the chair? Surely he would simply have to be able to make one of these attempts. Given the increasing dubiousness of the mental states as we travel down the regress of attempts, there are, I have suggested, only two relevant scenarios here: a normal agent - not subject to psychical constraint - must be able to attempt to rise from his chair if he is to be credited with the belief that he can so rise; an agent subject to psychical constraint must be able to attempt to attempt to rise from his chair if he is to be credited with that belief. (I would admit in these two orders of mental state, but no higher; however, the support I am offering the subjectivist does not turn on this.)

I cannot think of another candidate for the extra element turning belief that one is free into genuine freedom; I therefore conclude provisionally that the subjectivist wins the debate: if a rational agent has the cognitive and behavioural resources to earn a belief to the effect that he is free, then he must be free; nothing further is required for freedom, and there is no content to the supposition that he is illuded. What is more, it is likely that this

truth does not hold merely for the most part, but is true in every situation in which an agent believes he is confronting a choice between alternative courses of action.

What about the implication in the reverse direction? Does an agent have to believe he is free to be free? Here we can appeal to our earlier discussion of the spectator-agent. The hypothesized spectator-agent was not free, and we saw that he did not have the resources to construct A-series time. But an agent who did not believe himself to be free would have to conceive of himself as just such a spectator-agent. And that would undermine his ability to construct A-series time. In reality, of course, it would be very hard for an agent to think of himself as a spectator-agent. The reason for this is precisely that he would have to think time out the world - or at least out of the domain of action. Still, perhaps an agent could have the second-order belief that his first-order belief in freedom was an illusion. But to give content to the second-order belief, he would at least have to be committed to the coherence of the hypothesis of the spectator-agent; and that means he would have to be ready to live with the possibility that the world might be timeless. To suppose that a belief which you have is an illusion is to have some idea of what it would be to be disabused, even if you know you cannot in practice relinquish the belief. But it is quite implausible to suggest that we can give any content to the idea that the world might be timeless.

What this shows is that a rational agent who takes himself to act in time cannot but believe himself to be free. We get the implication from freedom to belief in freedom by noting that the

existence of a free transcendental agent is sufficient for the rationality of that agent, and sufficient also for the temporality on the way the agent thinks of his actions. The first of these steps was accepted in section 5: that freedom in action requires rationality. (And this must amount to self-conscious rationality, for an agent must have the ability self-consciously to deliberate about his actions. Rationality in action requires the availability of the reasons for action to the agent.) The second step was also argued for in section 5: a rational agent who acts freely in time cannot but conceive of himself as acting in A-series time, on pain of assimilating the past to the future, and thereby undermining the possibility of his own forwards action. I conclude therefore that a rational agent's belief that he is free is necessary and sufficient for his freedom. This is correctly labelled a subjectivist conclusion, but the subjectivity in question is, as we have seen, not such as to militate against the objectivity of the experienced item, namely freedom, at least not unless we are prepared to say that the world is, in itself, timeless. And that is a very hard saying to agree to.

3.7 Freedom and Internalism

Time is transcendental in the sense that there is no explaining why one dated time rather than another is now (3.4); freedom is transcendental in the sense that it fulfills no explanatory role in accounting for action (3.5). There must, as I have said, be some connection between these transcendences, and surely they must both also be connected with a further transcendence present in selfhood - that of the subjective features of experience. Colour experience, for example, is associated with the having of a subjective phenomenology: the subjectivity of the phenomenology is shown in the kinds of dissociation between phenomenology and functional role (and brain state) which appear to be conceivable.⁴⁰ Since functional role is what interlocks with explanation of action, it follows that phenomenology transcends explanation. In accounting for behaviour based on discrimination of colour, no appeal is made to anything beyond the discriminatory abilities themselves: to be colour-blind is to lack certain discriminatory abilities: the look of a colour is neither here nor there. I merely assert that there must be some connections between these transcendences; exactly how to draw them I am uncertain. It is unclear how illuminating it is to trade on the common element of transcendence; but I have nothing else at present to offer.

Kant argued that there exists a certain antinomy between the freedom of the transcendental self, and the conditioned nature of the empirical self. The position I have reached does not look quite like that, because on the empirical side I have not subscribed to physical determinism, but rather to the explanatory redundancy of

freedom. So it might seem at first glance that I have avoided the embarrassment of an antinomy. But in fact the situation is every bit as bad as Kant's (not that Kant clearly saw how bad his position was): we have reached a position from which it is an easy matter to generate contradictions.

Suppose that I have a choice between performing A or B. Suppose that I freely (and hence rationally) choose to perform A. Given the rationality of my action, I had some reason R, which I (intelligibly) judged to support the performance of A. But the freedom of my choice requires that, had I performed B, I would also have performed it freely, hence rationally, and hence I would have had another reason R', which I would (intelligibly) have judged to support the performance of B. But now R both cannot and must be stronger for A than R' is for B (and vice versa). R cannot be stronger, because A would then have been more rational and hence B would not have been genuinely free after all (on the principle that the freer an action, the more rational it is). Hence even my choice of A would not have been genuinely free, because the freedom of choice of one course of action requires the freedom, in counterfactual circumstances, of the choice of alternatives. By parity of reasoning, R' cannot be stronger than R. So R and R' must be exactly as strong as each other. But then my choice of A over B is inexplicable, and so irrational (or at least not fully rational), and so not genuinely free. So R must be stronger than R'. We arrive at the conclusion that action is genuinely free if and only if it is not genuinely free. Alternatively, to accommodate the fact that we are dealing with matters of degree, perhaps we ought to put the

contradiction like this: the freer an action, the more rational it is; but also the less explicable it is (because the freer and hence the more rational alternatives to it), so the less rational it is.

This contradiction would be avoidable if we could rule that an action may be free even if its alternatives would not have been. In that case, we could envisage A being more strongly supported by R than B is by R': freedom and rationality would coincide, without leading to the undermining effect run through above. But it does not seem to me that we can make the ruling. How can a choice be free if there are no alternative choices (as these would not be, if were the agent to act differently he would not be acting freely)? It really seems to be an a priori truth about free action that if an agent performs an action freely, then had he performed an alternative action, he would also have performed it freely.

The above contradiction is not the only one we must face. The need for freedom is guaranteed by the directionality of time; the need to dispense with freedom is guaranteed by the explicability of human action. Not only is human action in large measure explicable, and explicable to the point of enabling prediction to achieve a high degree of success; it is also the case, for reasons which Davidson has made familiar (in papers collected in his [2]), that human (or other) action must be explicable if it is to earn the title of action. A type of 'behaviour' which simply could not be understood as the intelligible upshot of (centrally) beliefs and desires intelligibly ascribable to the behaver would simply not be action. These two commitments when conjoined yield a contradiction, because the inclusion of freedom brings with it the inexplicability of free

actions. The rationality of human action, and its associated predictability and explicability, require that freedom be excluded from the components of action.

Freedom and the explanation of human action do not sit well together (as Kant observed: A 557/B 585); and, as we have seen, the mere assumption of freedom, together with some plausible constraints on the nature of free action, generates its own internal contradictions. For freedom both requires and inhibits the rationality (and hence explicability) of free action; and that rationality inhibits freedom. So freedom in effect inhibits itself: it is the incorporation of freedom into our comprehension of human action which creates the difficulties, rather than the rationality and explicability of human action. But we should be clear about the status of these contradictions. Until and unless a specific mistake is found in the transcendental deduction of freedom offered in this chapter, it will not be open to us to construe the derivations of contradictions as a reductio on the assumption of freedom. It does not seem to me fantastic that the contradictions should be simply ineluctable. At any rate I am not prepared to regard a contradictory outcome as worse than removing time from the world. (My own feeling is that the thesis that freedom constitutes the time-direction has such an intuitive 'rightness' about it, that some correct deduction of the thesis must exist, even if my own is faulty.)

The contradictions are surely the counterpart in the domain of agency of the inconcinnity between manifest and scientific images which we reached in the domain of experience. I do not see any way of resolving either of these difficulties, because I do not see any

prospect of abandoning the commitments on either side of each of the divides. The place of mind in nature - not only the scientist's nature but also manifest (human) nature - is, as it seems to me at the moment, an inexorable mystery. But it is at least possible, I think, to unify the mystery. I noted in chapter 1 that the internalist's claim could only be formulated in A-series terms: adopting his first-personal point of view, the internalist might claim, for example, that he could not fail to be conscious now. I suggested that the manifest image incorporates something like this internalism. Since now this internalism rests on the availability of the A-series, and the A-series is in turn constituted by the freedom of the transcendental self, it follows that this freedom is at the very core of the manifest image of man. Freedom is what makes possible both the subject's conception of his special access to his own experience, and the agent's conception of himself as acting in time; equally, freedom is responsible both for the difficulties involved in reconciling the manifest and scientific images of man (chapter 1), and for the difficulties integral to the manifest image itself (this chapter).

Indeed we can strengthen this position. Not only does internalism depend on freedom (via the availability of A-time), it is also the case that freedom depends on internalism. This is because the capacity of freedom to constitute the anisotropy of time depends crucially upon the free agent's ability himself to supply a guarantee that he is free; no one else can make the certification on his behalf. But the certification must be made, on pain of failing to deflect the scepticism which we saw was ruinous to the chances of

the compatibilist's agent of directing time. Hence the deduction of freedom on the basis of time depended on internalism. Let us see how this works in more detail.

In our discussion of the spectator-agent's inability to direct his own time, we saw that the spectator-agent could not rule out the possibility that he was being manipulated - by being fed with quasi-precognitions as well as memories, between which he had no means of distinguishing - and so did not have the resources out of which to make a non-arbitrary distinction between the past and the future. But the free agent is not liable to such a scepticism: he can always assure himself that he is not a spectator-agent, by allaying the suspicion that impressions which he takes to be his memories are in fact quasi-precognitions. He does this by acting so as to falsify the hypothesis that they are precognitions. And we saw that this action in turn depended on his ability to make a decision lying beyond the compatibilist's purview (a decision such as the compatibilist's agent cannot make). In this way the free agent can assure himself that he is free. Now what we need here is the further (internalist) conclusion that only he can provide this assurance.

But it is surely clear that this further condition must hold. For the scepticism which ruined the spectator-agent's ability to direct his own time must generate a related scepticism in anyone who, adopting the third-personal stance and without proceeding by directing his own time first, attempts to direct the time of another agent. The related scepticism is this: how can he assure himself that he is not in the presence of a spectator-agent? There is no guarantee, from anything which a merely third-personal view

discerns, that he is dealing with a free agent. As far as the provenance of a man's actions go, I have already argued that the explanatory and interpretative project proprietary to the third-personal stance discerns no such freedom - indeed positively excludes it. Schopenhauer said (p.42) that we treat each other deterministically: it is only in respect of his own actions that an agent has to make room for freedom. I have subscribed to the analogous principle that we treat each other as rational beings, and that rationality as discerned from the third-personal stance (i.e. rationality interlocking with explanation and prediction) excludes freedom. The externalist's stance cannot understand its subjects if it does not treat them as rational beings who behave explicably and predictably; so, since freedom introduces inexplicability into human action and is a bar to predictability, the externalist cannot discern freedom. (If, as many would say, to treat a man as a moral being is inevitably to regard him as free, that treatment itself then stands in need of a warrant: it cannot certify itself.)

The way in which freedom constitutes time must now involve internalism, because it involves the free agent being in a position to decide, incompatibilistically, to act in such a way as to falsify the hypothesis that the impression he is currently enjoying is a quasi-precognition; and the certification of his decision as a decision can only be supplied from within, because the decision only reveals itself as such to the internal view. The outside interpreting observer - and so the externalist - cannot see it; he is in exactly the same position as the compatibilist: as far as the externalist is concerned, the agent might be a spectator-agent,

someone who at some suitably deep level does not choose what he does. It is of the essence of the ability to make the free decision which constitutes the anisotropy of time that it is such that the internalist's account of it is correct. The metaphysical deduction of freedom on the basis of time presupposes the epistemological doctrine of internalism: that the subject has authority on the question of (in this case) when he makes a decision, and the nature of the decision he then makes.

Taking the conclusions of the first chapter and this chapter together, we can deduce, on the assumption of time's anisotropy, a logical structure with mutual implications between internalism, the ineliminability and transcendence of A-time, incompatibilist freedom, and the existence of the free transcendental self. The structure as a whole, I suggest, encapsulates the manifest image of man as subject of experience and as agent; further, it yields both the difficulties and contradictions we have encountered. It would be neat if we could construe the derivation of contradictions as a reductio and contrapose, thus getting rid of all our difficulties at one fell swoop. But, as I hope to have made clear, the price of that manoeuvre is too high. It would entail not only an implausible externalist construal of experiential consciousness, but (what is worse) a commitment to the timelessness of the world. We can perhaps just see how an externalist view of consciousness might be able to make good. But we cannot live with the possibility that the world might really be timeless: that would undermine everything. The mystery engendered by on the one hand our difficulties in reconciling the manifest and scientific images of man, and on the

other the contradictions within the manifest image's commitment to freedom, is preferable to that. (Even within the manifest image, there is a tension between first- and third-personal perspectives; but neither is available without the other.)

In the sequel, I wish to explore two topics connected with the argumentation of this chapter. Firstly, I shall try to argue for a crucial distinction between space and time: that space, unlike time, is not transcendental. Secondly, I shall explore in more detail the notion of the transcendental self which we have arrived at. Among other ways in which the self has commitments both to the transcendental and to the empirical, I shall have to consider how it is possible for that self to inhabit both a transcendental temporal order and an empirical spatial order.

Chapter 4: Experience and Space

4.1 Particularity and its Concomitants.

As free agents requiring things to manipulate, and as subjects of experience requiring things to perceive, we inhabit a world of material objects and their properties. These objects are partly distinguished by their possession of spatiotemporal location. I shall not attempt to delimit them further, or even to discriminate them from other objects, such as events or processes, which might be said to have spatiotemporal location, but rather proceed with an intuitive understanding of the sort of thing I mean: I shall be concerned in this chapter with such material or concrete objects as tables and chairs. I shall ask the question: what is needed to constitute the spatiality of such objects, and consequently of the spatial order itself (which, I shall take it, following Leibniz, simply consists in the spatial relations of such objects)? I shall try to show that the transcendental self is not competent to discharge this role, and that the reason for this is that space, unlike time, is not transcendental. Given the main result of the previous chapter, that the free transcendental agent is needed to constitute the temporality of the world, a sharp wedge will be driven between time and space. Taking the last chapter and this together, I hope to provide some embellishment of the Kantian division between time as the form of inner sense and space as the form of outer sense.

Of course spatiotemporal or spatial objects (the two categories are in fact co-extensive) are not the only sort of object: there are

also, for example, universals (which have neither spatial nor temporal location), and transcendental selves (which have temporal but, as we shall see, not spatial location); but they are an extremely important class of object (other sorts of object being, so to speak, made for them).¹ In the rest of this chapter I shall simply concern myself with spatiotemporal objects, and especially with material objects. I shall where convenient expedite the discussion by speaking about Objects - meaning thereby material objects - and capitalizing to register the restriction. Material objects, then, or Objects, have their properties. Such properties do not attach to their bearers as coats hang on a peg. Pegs persist through changes of coat hung on them, and they persist through total such change. They may also exist without ever having coats hung on them. Material objects, on the other hand, while they may persist through changes of property, may not persist through total such change. And they may not persist during a period in which no properties inhere in them. They do not have a transcendental core. Pegs have natures which are unaffected by whether or not, if ever, coats are hung on them. But the nature of an Object is precisely specified when its properties are specified. There is accordingly no such thing as a bare Object: an Object cannot be conceived of as stripped of and existing in independence from its properties. Remove all the properties from an Object, and it thereupon ceases to exist.

This is a familiar line of thought. What it shows is that to talk of Objects is not to prescind from a commitment to the existence of properties. Nor is it of course to talk merely of properties, since properties are universals. In fact, to talk of

Objects is to talk of instantiated properties. We do not regard Objects as in any sense fictions: we cannot make do without instantiation, for that would be to make do without the particularity of the things we daily encounter. Nor can we regard Objects as constructed out of non-spatiotemporal properties in special combinations. Abstract objects are just such combinations; but if we tried to construe material objects along the same lines we would collapse the distinction between an Object and its counterpart in a symmetrical universe. But there surely is a distinction here: we are surely correct to suppose that we can distinguish numerically, as we say, between Objects which are qualitatively identical.² Leibniz must have been wrong to suppose that two Objects cannot differ merely in their spatial relations to other Objects. (I extrude temporality from consideration, since it is uncontroversial that two otherwise identical Objects can differ merely in point of temporal location.) In the familiar example of the chessboard,³ we can distinguish one square from a qualitatively identical counterpart obtained by reflection in some axis. We do so in terms of the different spatial locations of the two squares.

To make this distinction is not to import a Newtonian conception of absolute space: the distinct places occupied by the distinct squares are indeed individuated by spatial relations to other things, and not by position in absolute space. But it is crucial here that the spatial relations, by means of which we individuate places, incorporate essential reference to things, that is to say, particulars.⁴ A place is individuated by its spatial relations to particulars, and there is no prospect of eliminating

the occurrence of particulars in this constitution; otherwise there would be nothing to figure as the relata of the relations purporting to individuate places. It follows that we must think in terms of a mutual interdependence between particularity and spatial location: the possibility of the one requires and sustains the possibility of the other. There may be a sense of the phrase 'spatial location' which does not involve itself analytically with particularity (and so is committed to purely general characterizations); but I shall not be concerned with any such sense. Particularity - in the sense of numerical particularity, i.e. the sense which allows us to say that an Object is a different particular from its counterpart in a symmetrical universe - particularity in this sense seems to be fundamental to our understanding of the world.

The metaphysician cannot stop there, however. He must ask at least two further questions: firstly, is particularity as defined in the previous paragraph (which I shall henceforth just call 'particularity') merely a contingent feature of our world-view, or is it a necessary feature of any world-view? Secondly, what is the justification, if any, of our belief in the existence of such particularity? (Is that belief dispensable?) These two questions are not unrelated. To show that particularity necessarily attached to any world containing Objects (there is no circularity here since Objects have been defined by paradigm) would be to justify our belief in it; and to justify that belief would plausibly amount to (because would require nothing less than) showing that particularity necessarily inhered in any world of Objects. I shall assume that the two questions are both aspects of a single question.

I want to suggest that the justification can be carried out, but before giving an argument to this, I need to specify some more of the metaphysical and epistemological commitments of particularity. I start with the observation that the mental correlate of particularity has to be de re or genuine singular thought, that is, thought without a descriptive surrogate. Now a genuine singular thought about an Object is a thought which bears a specially intimate relation to that Object, such that if there were no Object of the relevant sort, there would simply be no thought of that sort to be had. It follows that the content of genuine singular thought cannot be specified descriptively, since the availability of descriptive thought is in general indifferent to the existence or otherwise of an Object satisfying the description. Typically, thoughts of the form 'that F is ...', and 'a is ...' (where 'a' is a proper name which does not itself abbreviate a description) are genuine singular thoughts. I shall not elaborate on this point here, since it has been thoroughly discussed in the literature.⁵ The epistemological correlate of genuine singular thought is acquaintance. Following Russell,⁶ we may say that a mind which essays a genuine singular thought is, if successful, acquainted with the Object of the thought. This contrasts with the descriptive knowledge of an Object which a man may have in entertaining a general thought of the form 'the F is G'.⁷

So far I have not offered an argument for the connection between genuine singular thought and particularity; I have just given a sketch of the view I believe to be correct. I shall now make good that deficiency. Consider first the implication from genuine

singular thought to particularity. Surely, it might be objected, there cannot be a dependence of the former on the latter, since even if there were no particularity, so no principle on the basis of which to distinguish a world from a symmetrical duplicate of it, genuine singular thought would not thereby be undermined: we would still in fact frequently think of Objects in the de re mode, or non-descriptively. But the objection is misconceived. If we take proper account of what is involved in the supposition of a world without particularity (i.e. a world indistinguishable in principle from a duplicate and hence entirely capturable in the net of descriptive thought), we see that in that world there can be no genuine singular thought; indeed there cannot even be purported singular thought (singular thought pro tempore), for in a world without particulars what would the text of such a thought be? Such thought would be aiming to do something which ex hypothesi cannot be done, namely pick out an object from its counterpart in a symmetrical universe. But by assumption we cannot distinguish these^e universes, so that we cannot even say what genuine singular thought would be trying to do if it did.

Let us now turn to the other side of the equivalence: the question whether there is an implication from particularity to (the availability of) genuine singular thought. Here the argument makes use of the supposition of massive duplication (a supposition which the assumption of the existence of particularity entitles us to), that is, the supposition that there could be a universe symmetrical with our own. If there were such a universe, and if all thought were descriptive, there could be no picking out of the non-abstract

objects in our world. For no matter how full one's description of an object, there would always be at least two objects satisfying it. In this situation, we need a species of thought which goes beyond the merely descriptive. But it might be objected: if the universes really are completely symmetrical (and if they weren't, individuating descriptions could be constructed), what would be the point of genuine singular thought? What could it achieve that descriptive thought could not achieve? Surely all we actually need is thought which individuates up to total descriptive isomorphism? And if we don't need to go further, there is some plausibility in the suggestion that we can't: how could meaning outstrip any conceivable use?

This line of thought takes its cue from consideration of thought about abstract objects, and to rebut it we need to call to mind the difference between abstract and non-abstract worlds. We should first of all note that there is no possibility of constructing an argument from massive duplication in the abstract case: and that is because there is no such thing as genuine singular thought about abstract objects. All thought about abstract objects is descriptive, because abstract objects have purely qualitative identity conditions: it is not possible to have qualitatively identical but numerically distinct abstract objects. This in turn is grounded in the nature of the space - logical space - inhabited by these objects. Position in logical space is determined purely qualitatively. There is no semantic way to distinguish isomorphic structures, only syntactic.

Turning now to the non-abstract case, the argument I offer

proceeds by spelling out exactly what is involved in the assumption that the symmetrical universe move makes sense (that it is possible to distinguish numerically between qualitatively identical objects). One can imagine two qualitatively identical tennis balls, say, sitting side by side. They are different balls. Their non-identity is not constituted by the different general spatial relations they bear to surrounding objects: one can think that surround away, or one can locate the balls at equal distances from and on either side of the axis of reflection dividing the universe from its symmetrical duplicate (so that the one ball is the counterpart of the other), and the balls do not then coalesce. How is it possible for thought to concern one tennis ball rather than its counterpart, given their numerical distinctness? Consider an attempt at a genuine singular thought; suppose that on some occasion I think

(U) That tennis ball is white.

What makes it the case that I am thinking about one tennis ball rather than the other? The answer is that the demonstrative component of the thought is conceptually simple: it is not possible to think of two differently located Objects both as 'that Object'. If I am able to think of one of the balls as that ball - and this is what I am attempting to do - then I am simultaneously unable to use the same thought-episode to think of the other tennis ball, given its distinct spatial location. We can distinguish my thought about that ball from any thought I might have about the other ball, because we can distinguish the places they concern, and because the demonstrative content of each thought relates to its peculiar place; it follows that a thought-episode with some such demonstrative

content cannot relate to more than one place.

In effect we are appealing here to an essential feature of genuine singular thoughts: they are had by subjects enjoying a point of view on the world. Of course if the universes are genuinely symmetrical, there will not merely be two tennis balls in play, but also a counterpart of me simultaneously entertaining a thought which he can express by uttering (U) on his side of the axis. But that fact does not yet disturb my ability (and his) to entertain a thought about just one of these balls: we are given that I am not (numerically) identical to my counterpart, so that our respective acts of thinking are non-identical. And this in turn gives us what we need to assure ourselves that each thought concerns the place in its universe and not the other. For my thought about the place where that ball is has a spatial content which is essentially egocentric (that is what makes it an attempt at a genuine singular thought): I think of the place as, say, over there at about three feet from me. Or perhaps, I just think of it as here. But now as Evans says ([2], chapter 6), it is not possible to think of two different places simultaneously as here. (It is not possible to think of two places as over there at about three feet from me.) So I could only be thinking of the parallel place in the symmetrical universe, as well as that place in my universe, if the thought-episodes of me and my counterpart were numerically identical - in effect if I and my counterpart were numerically identical. And that we are assuming not to be the case. The important fact, then, is that my counterpart would not be me. There can be irreducibly singular thought about the non-abstract world because subjects of thought are irreducibly

singular.

On the assumption of the possibility of particularity, therefore, genuine singular thought is also possible: it is the mental correlate of particularity. We have reached the point where we can say that there is a tight triangle between particularity, spatial location and singular thought. We have argued for relations of mutual implication between the first and second of these items, and between the first and the third.⁸ But although this clarification of conceptual connections is certainly important to us, it does not immediately help us answer the two metaphysical questions we started with. We want to know what the metaphysical status of particularity is, and our entitlement to operate with it. We can now say that these questions concern not only particularity, but also the availability of spatial location and genuine singular thought; that is an important step forward, but we need to know now how the triad as a whole fares before the tribunal of justification. The question we must ask is: what, if anything, constitutes the triad?

4.2 The Constitution of Particularity.

Why should there not be such things as tables and chairs in a world without any of the elements of the triad? Such things would not be particulars (in the sense that there would be no principle of distinction between qualitatively identical Objects), they would not figure as the objects of genuine singular thoughts, and they would not be spatially located in the sense that they would not be located at places which were themselves particulars in the above sense. (Of course anyone taking this line is going to deny that places need be thought of as particulars: he will want to operate with a notion of general spatial relations.)

In the above discussion of the connection between the possibility of genuine singular thought and the distinctness of my counterpart and me, that numerical distinctness was an assumption of the argument for the availability of genuine singular thought. We were entitled to the assumption because we were asking after the availability of genuine singular thought in a world of which the assumption held, that is to say in a world housing particulars. The challenge to us now is to make good our entitlement to believe that assumption. Anyone who wanted to deny the necessity of the triad to a world of Objects would have to question its credentials. He would have to claim that I and my counterpart in a symmetrical universe were identical. Now if the challenger were right about this, it would follow not merely that the triad would not be a necessary feature of any world containing Objects, but more strongly that it

would necessarily be absent from such a world. For if I and my counterpart are identical, it follows - according to a plausible principle to the effect that identicals are necessarily identical - that no sense (no metaphysical sense) could be made of the hypothesis that we might not be identical; and thence it follows that no sense could be made of the hypothesis that there might be genuine singular thought (since my non-identity with my counterpart was integral to the possibility of genuine singular thought). In that case the obtaining of the triad as a whole would be an impossibility. So either the triad necessarily obtains in any world containing Objects, or it necessarily fails to obtain.

Now the status of the triad cannot be defended merely by appealing to the fact that the tennis balls in the envisaged scenario are in different places. This simply begs the question against the challenger, who will of course contend that the places in question are one and the same place: 'in a genuinely symmetrical universe,' he will say, 'no empirical purpose could be served by distinguishing one of the universes from its 'reflection'. For what sense can be attached to the distinction? Is it not like the supposition that God could have created the universe sooner than He did, on which Leibniz remarks that the supposition is chimerical?' But if my discussion of the transcendental status of time was on the right lines, we should reject Leibniz's remark. There would be no empirical difference between the actual world and the world if God had created it sooner. But there would be a transcendental difference: in the latter world, I would not now be free to do what, in the former (actual) world, I am free to do. The transcendental

status of this remark is indicated by the fact that if we substitute a date for the underlined occurrence of 'now' in the above sentence (so turning the claim into a B-statement), the claim becomes false. And that transcendental status should embolden us to see if we can say something similar about the spatial case. Clearly there need be no empirical difference between the two universes, but does it follow that there would be no difference at all?

Perhaps, it might be thought, a principle of distinction between the two universes could reside in their opposite spatial orientations. That is to say, given the incongruence between an oriented object (such as a glove) and its counterpart in the symmetrical universe, to assert the identity of the universes would be to assert the identity of (say) a left-handed glove and its right-handed counterpart in the 'reflected' universe. But the distinction in handedness of the gloves surely guarantees their numerical distinctness. In a similar case, I might stand looking at my counterpart across the axis of reflection of the universes, and think that my heart is on the left of my body; my counterpart, meanwhile, entertains the thought that his heart is on the right of his body. Are these thought-episodes not guaranteed numerical distinctness by the difference between left and right orientations?

The answer is that in the envisaged scenarios the distinction between left and right orientations does not apply in such a way as to differentiate either the gloves or the contents of my and my counterpart's thought-episodes. Certainly there is a distinction between left and right orientations, and certainly in one sense the symmetrical universe is left-right reversed, so that in that sense

the counterpart of left here is right there. But unfortunately these facts do not suffice to show that there is a non-identity between left here and right there. For it is plausible that the bare incongruence between left and right orientations as it applies within a universe is the only thing which is constitutive of the relation between them: there is no absolute left or right orientation. That bare incongruence obtains in the symmetrical universe as much as in this one; but there can be no bar on the basis of the present considerations to identifying left orientation here with right orientation there. What it means to say that there is no absolute left or right is just that the distinction between left and right is not sensitive to the distinction between a universe and its symmetrical duplicate. In crossing from one universe to the other, as it were, left becomes right.

The ^lposition is just like this: suppose that space in our universe is in fact non-oriented, so that if you move a left-handed glove sufficiently far in one direction it eventually comes back as a right-handed glove (just as figures such as Nerlich's Knees, although incongruous on a flat table-top, are congruous on the surface of a Möbius band).⁹ That is a physical possibility. If our universe were like that, there would be no absolute sense in which a glove was left-handed; nevertheless it is clear that the distinction between left and right orientations would not lose its applicability: a glove would not therefore be of neutral handedness. By engaging in space travel, gloves could change their handedness, but there would still be two sorts of handedness available.

Returning to the case of the universe and its symmetrical

counterpart, one can see that within each universe there are two sorts of orientation available, but that the orientations do not have inter-universe status. Moving from one universe to its reflection is like the whole universe engaging, per impossibile, in whatever space travel would be necessary to reverse orientation. When one glove undergoes such reversal, the possibility of its doing so depends on the region to which it returns not travelling likewise. No sense can be attached to the universe as a whole changing its orientation; so no sense can be attached to the supposition that the universe enjoys an opposite orientation to its 'reflection'. We should consequently say that - as far as the present considerations go - when I think of my heart as lying on the left of my body, and my counterpart thinks of his heart as lying on the right of his body, the thought-contents are one and the same. And then there still seems to be no objection to construing the thought-episodes (and me and my counterpart, and the universe and its 'reflection') as identical also.

There is no prospect of distinguishing the universes along these lines. Let us now ask the question: could bringing the universes into interaction with each other help, and in particular, could bringing me and my counterpart into confrontation lead to a principle of distinction between us? Consider then the following scenario: I approach the axis of 'reflection' between the universes. As I do so, of course, my counterpart does the symmetrical equivalent in his universe. We arrive at the boundary simultaneously and look across at one another. Now this way of putting it tendentiously presupposes the non-identity of me with my

counterpart. The challenge then recurs: what justifies this presupposition? Why not say there is just one universe here? Of course we have to say something about the content of my visual experience: I seem to see someone who has all my empirical attributes; we might then confront the challenger with the question: how can I literally be seeing myself? The challenger's response must be to construe the boundary between the universes as a mirror.

By way of approach to the interpretation which the challenger is putting in front of us, consider first the experience of looking at oneself in a mirror. What does one see? 'The eye sees not itself, but by reflection...'.¹⁰ Of course you can see the corporeal eye, when you look in a mirror, but you cannot see the seeing eye; you cannot see yourself seeing. (Or one might put it: you can see your body but not your transcendental self (cf. TLP 5.633).) This is indicated by the following fact: if you did not know what you looked like, then in a situation in which you were confronted (perhaps by the use of trick mirrors) with two appropriately positioned images of figures looking at you, you would not be able to tell which of the figures, if either, was you (i.e. your body). If, per impossibile, you could turn the trick Hume notoriously attempted, and observe your own seeing soul, then you would be able to tell. When you observe yourself, as we say, in a mirror, a two-stage procedure occurs whereby extraneous knowledge of what empirically you look like combines with your knowledge of your orientation, and a simple grasp of the physical properties of mirrors, to inform you that you are looking at yourself (your own body). You do not observe yourself (i.e. you the subject of your observation) in a mirror, and

hence even your knowledge that it is your own body that you see there is essentially fallible. This contrasts with the status of your knowledge that you are the subject of your experience. In general, of any thought you have it is infallibly epistemically available to you that you are the thinker of it. This is the core of Descartes' cogito argument. (I return to this in the next chapter.)

Suppose I can observe my counterpart's body across the boundary between the universes at what appears to be a distance of two yards. Let us suppose that I reach out and point to his body, he simultaneously gesturing towards mine. How many demonstrations are there here? Suppose we say, with the challenger, that there is only one. Then is something demonstrated, or nothing? It is surely counterintuitive to say that nothing is demonstrated: ex hypothesi I see my counterpart, he sees me, and we attempt to demonstrate one another. Even if there is only one demonstration, it surely has an object: in that case, the object must be my body (=my counterpart's body). We might now be tempted by the following line of thought. I aim to demonstrate something at a certain distance in front of me. But my body (i.e. the place where I am seeing from) is not, and cannot be, at any distance from the origin of my demonstration: that would be for my body to be at a certain distance from itself. So I cannot be demonstrating my own body. But I am at least demonstrating my counterpart's body. So I and my counterpart, pace the challenger, are non-identical.

Unfortunately, however, this argument proves too much; it would make it impossible wittingly to point to one's own body (or

anything, for that matter) in a mirror. And that surely is possible, so long as one is undeceived about the operation of mirrors. Still, the proviso is important: it can only be in a derivative or non-basic sense that I am pointing to the place where the reflected person stands, rather than to the place behind the mirror where his virtual image looks as if it is located. In the basic case, there must be coincidence between the region where one's finger is targetted, and the location of the demonstrated object. (The point of the basic/non-basic terminology here is to capture the dependence of non-basic demonstrations on the availability of a practice of making basic demonstrations.) Perhaps the cost of accepting the challenger's claim that I am identical with my counterpart is that the demonstration in question be construed as non-basic.

But how much power does this point have? Suppose that in our envisaged scenario I do not take myself to be making a non-basic demonstration. I take myself to be demonstrating an object two yards away from me. As far as I am concerned the situation presents itself as a basic case does within my universe. Given this intention, it may indeed seem plausible to say that I cannot be demonstrating my own body. For surely in cases of such mismatch between referential intentions and the facts, the attempted thought is not well-grounded, so that the resultant thought (if any) is not of just the type essayed. In the present case too, if the only sense that could be made of my referential act entailed saying that I demonstrated myself (non-basically), would that not be a rationalization, since would it not be arrived at by disregarding my intention?

The challenger can reply here that it is open to question how damaging to his position he ought to regard such mismatch between intention and interpretation. For it is not as if the failure of my intention need indicate, contra the challenger, that there are, after all, two Objects in play. Perhaps I really do in this case essay a thought which is unavailable to me. My essays are not self-fulfilling. The challenger claims that my demonstration of my counterpart's body is identical with his demonstration of mine. As a duple person (so to speak, and to put it paradoxically) we point at both bodies; we control both bodies and act from them. And then the challenger's claim is just that when the scenario is properly assessed (when the paradox is removed) the duplication vanishes. There is in reality only one person here, and only one universe. The act of demonstration here has to be construed as pointing at one's own body in a mirror, but in a slightly peculiar sense. To put it paradoxically again, a man is arguably not sufficiently singular to refer to a non-distinct counterpart in a symmetrical universe, because there is no possibility of such counterparts and such universes. A man is not as singular as that. We do not want to say that in the envisaged scenario I refer to nothing; the challenger's story could be that (in the non-basic sense) I refer to my own body.

But the deeper riposte open to us is to question the way the basic/non-basic distinction is being applied in the present case. We can present both the challenger and his opponent with a dilemma. Either the symmetrical universe is logically cut off from my field of action or it isn't. Let us first suppose that it is - i.e. that it is logically impossible for me to move to the position two yards

ahead of me which I am attempting to demonstrate - then the demonstration, basic or otherwise, must also be logically unavailable to me. In that case two consequences will ensue: we shall only be able to construe the boundary as analogous to a mirror, and any demonstration I make will relate to my own body, and basically so. As regards the first of these consequences, we must say that the position two yards 'ahead' of me and supposedly beyond the boundary does not exist. If I try to go there, then (let us say) I end up back at my starting point. I could not literally just hit a physical barrier, since then it would not be logically impossible (contrary to hypothesis) for me to cross to the other side of it. Or again, it is as if (but only as if) space curves back on itself at this point; and since it is logical space and not physical space which is curving back, no sense can be attached to the idea of a position beyond the 'boundary' or the 'turning point'.

As regards the second consequence, the fact is that if I am logically cut off from the symmetrical universe, there can be no warrant for supposing that if I gesture towards my counterpart I can at best be demonstrating myself in a non-basic sense. If it were the case that any action I could take in the direction of my counterpart in fact rebounded on myself, we should have to say that in demonstrating my counterpart I would be demonstrating myself basically; for the basic target of a demonstration is just that region into which I would act if I were following up my demonstration. In thinking of the position two yards ahead of me I am in fact thinking of the position I am at: for that position is,

at this point in the universe, two yards ahead of me. The paradoxical-sounding nature of these assertions reflects the fact that we have stumbled on a crucial incoherence in the dialectic, and one which (on this horn of the dilemma) favours the challenger: we are trying to describe, in terms of the original scenario (i.e. pictorially) a situation in which the universe and its duplicate are logical replicas of each other. But no such duplication could be spatially reproduced, because spatially located distinct objects always have the logical possibility of interacting with each other, and of moving to any spatial position. So if I could actually see my counterpart two yards away from me, the bare logical possibility of moving to the place where he seems to be (whether or not there is a physical possibility of doing so) would obtain. Contraposing, since on the present assumption I am logically debarred from acting into the place which seems to be two yards ahead of me, it follows that the terms of the original scenario - my seeing a counterpart across a divide - are not on this assumption coherent. The challenger's construal of the boundary as a mirror does not work, but that is because he is right about the identity of the universes. On this horn of the dilemma, the challenger is right to conflate me and my counterpart, and in general the universe and its duplicate.

Let us now move to the other horn, and suppose that the symmetrical universe is not logically sealed off from my influence. In that case I can move to the position occupied by my counterpart, and he can move to my position (we can smash the mirror, as it were, and walk through it). But now what makes it the case here that I have not merely doubled back on myself (that space does not

logically curve through 180 degrees at this point)? Surely only that I and my counterpart can act differently from one another; i.e. that it is logically possible for an empirical difference to emerge between us. On this horn of the dilemma, the challenger loses. Even if a universe and its duplicate are empirically identical (so that no description can be concocted which truly applies to one and not the other), as long as there is the possibility of an empirical difference developing - a difference capturable descriptively - then it will be possible to distinguish between the universes, and it will be right to say that there are two universes in question, not one. In that eventuality, the triad will be legitimately available to distinguish Objects in one universe from their counterparts in another; and, indeed, as long as no de facto difference emerges between the universes, nothing else will be competent to discharge that office. (Of course the triad does not and cannot serve to distinguish a universe from its logical duplicate; that is because there is no distinction to be made here.) It should be observed that while the triad is itself constituted by the possibility of descriptively capturable difference between the duplicate universes, genuine singular thought does not abbreviate, or go proxy for, descriptive thought; it could not do so, since de facto duplication is always a possibility: but there is no doubt, when I think or talk about something in the de re mode, that I am thinking or talking about something in my environment, and not in the environment of any counterpart of mine in a duplicate universe (cf. the argument of 4.1).

The justification of our practice of operating with the triad

of particularity, genuine singular thought and spatial location lies, then, in this: that as between two Objects which are qualitatively arbitrarily similar the possibility of some qualitative difference developing obtains. The universe and its duplicate may diverge empirically; or at any rate if they may, that is the principle of distinction between them. If they cannot, there is no distinction between them, and the challenger is right to conflate them. This is the upshot of reflection on the dilemma. That the above principle of the distinction between two qualitatively identical Objects is a necessary truth about Objects, and not just a contingent feature of our world, is shown by the fact that Objects, as we said at the beginning of this chapter, are purely empirical: they do not contain a transcendental core (see further in justification of this claim 5.3). Clearly, if all of an Object's properties are empirical, then it will be distinct from another Object just if it does or can differ from that other Object empirically, and not otherwise; whereas on the assumption that the universe and its replica are logical duplicates, the conceivability of a distinction between an Object (such as my body) and its counterpart breaks down.

Spatiality, unlike temporality, is not then constituted by anything transcendental, and so a fortiori not by the transcendental self. Indeed it could not be so. For without at this stage inquiring into the identity-criteria of transcendental selves (a task for 5.3), we can nevertheless assert this much: it must be possible for two transcendental selves (say, I and my counterpart in a symmetrical universe), to act differently (empirically differently)

if we are to earn non-identity with one another. I must be able to sit down without his having to do so too. If I could not act differently from my counterpart (in logical analogy with my physical inability to act differently from my reflection in a mirror), that would go to show that we were not really two persons after all, but one. So the distinctness of a universe from its duplicate is not constituted by my (prior) distinctness from my counterpart in it; rather, both of these non-identities are dependent for their constitution on the further condition that the universes in question may take separate empirical paths, that they may differ qualitatively from each other somewhere.

Space is a fundamentally empirical phenomenon; time, transcendental. In a sense this is already obvious from the difference between 'here' and 'now'. There is no absolute here: what makes a place here for a thinker is its proximity to that thinker's spatial position. But his spatial position is in turn not absolute, but is itself constituted by the spatial relations obtaining between the thinker's body and other Objects. On the other hand, we have seen that the nowness of the present moment is an absolute matter. Although there is equality between moments in the sense that every moment is, in its time, present; still there is an absolute, and transcendental, sense in which, for example, twelve noon of 26th March 1988 is (as I write this) now, and is so for every thinker. The challenger was wrong, as I pointed out (4.1), to suggest that no sense could be made of the hypothesis that God might have created the universe ten minutes earlier; we can make transcendental (though not empirical) sense of this hypothesis. But it turns out that he is

quite right in his parallel claim about space. We can make no sense of the supposition that God might have located the whole universe three feet to the left. The universe as a whole is not located anywhere at all in space, and so a fortiori could not have been located anywhere else. To move an object three feet to the left is only possible if a sufficient number of the surrounding objects stay put; to move everything three feet to the left is accordingly not a possibility. Let me repeat that this claim, and the argument of this chapter, is not supposed to be a move in the dispute between Newton and Leibniz over the question whether space is absolute or not. I have already said that I accept Leibniz's view that space is entirely constituted by spatial relations. The question for me has been: what constitutes those spatial relations (and hence the triad as a whole)? The answer is that they are not constituted transcendentally, but empirically.

Of course it remains the case that, as I have said, de facto duplication is a possibility. The possibility of empirical difference between duplicates supports the very possibility of duplication; but clearly where duplication occurs, the empirical facts will underdetermine the difference between a universe (or a place within it) and its duplicate. Now we do not want to express this underdetermination by saying that space is transcendental, because there is no guarantee that the empirical facts will, in any given case, underdetermine the difference between a universe and its duplicate. One might say: the underdetermination does not go very deep. But there is a guarantee that the empirical facts will underdetermine which moment is now: there could never be an

empirical specification of what it is in virtue of which a moment is now (3.4). We can express this distinction (which is reflected in the relativity of 'here' and the absoluteness of 'now') by saying that space is empirical and time transcendental.

Our conclusion sits well with Kant's distinction between time as the form of inner sense - i.e. the order in which the subject apperceives his self and his representations - and space as the form of outer sense - i.e. the order in which the subject intuits the objects of his representational experiences. On Kant's picture, time is closer to the self than space; we have been approaching the topic from a different direction, but we have arrived at a similar conclusion. It was indeed clear from chapter 2 that the subject is not confined to just one spatial option: we vindicated the possibility of at least one quasi-spatial alternative. But obviously no such range of alternative time-orders is available to a subject. In fact we have made the distinction between time and space go deeper: time is closer than space to the self because it is the order of the self's activity (that activity constitutes time), whereas space is not constituted by anything the transcendental self does, but by the empirical relations between the objects which it houses. Time, being transcendental, needs something transcendental to constitute it; space, being empirical, is appropriately constituted empirically. I pointed out in 4.1 that there can be irreducibly singular thought because I am irreducibly singular. We can now gloss this as follows: it is because my body is irreducibly singular that I can enjoy singular thought, and - what goes with this - representational experience of the world. The singularity of

my body is not then in turn constituted by my singularity as transcendental self, but by the empirical relations which it enjoys with other empirically constituted items. I am singular only up to (and not including) logical isomorphism with a duplicate self in a duplicate universe. As subject of experience, I am constituted empirically by an anterior empirical spatial order: that is a major conclusion of this chapter. It contrasts with the conclusion of the last chapter, that in the case of agency the direction of determination is the other way round: the temporal order is constituted transcendently by the prior transcendental status of the free self. The transcendental self is not as such spatially located: it achieves spatial location when it enters a body.

How do these aspects of the self cohere? We can make a start on an answer by noting that it is only in one sense that time is closer than space to the self. In another sense, the agent's activity is necessarily located in some space or other, as the locus of his activity. The transcendental cannot simply dissociate itself from the empirical: it must be realised in an order which is not itself transcendental. These remarks about the necessary engagement of time with space (through the activity of the free self) precisely parallel remarks which we must make concerning the essential possibility of embodiment of the transcendental self. It is to this issue, and others relating to the nature of the self, that I finally turn.

Chapter 5: The Transcendental Self.

5.1 The Phenomenology of the Self.

In chapter 3 we deduced the existence of the transcendental agent as the basis of the temporal order (with freedom occurring as the middle term in this deduction); in chapter 4 we saw that the transcendental is neither needed nor competent to constitute the spatial order, because the spatial order is purely empirical. But we have also noted that the transcendental and the empirical orders are connected in this way: the transcendental self, whose activity as agent constitutes the temporal order, needs some space in which to conduct its activity. And that space will necessarily be constituted spatially by the network of spatial relations which comprise it. That empirical order will then be available and competent to constitute the so far transcendental agent as a subject of representational experience, because it will only be in virtue of the empirically constituted particularity of Objects in the world that the subject will be able to entertain genuine singular thoughts - and hence enjoy representational experience - in which those Objects figure as intentional objects. Hence a self which, as it were, started life as an agent would have what it takes to acquire subjecthood in addition.

Contrariwise, while considerations of subjecthood do not similarly compel a deduction of a transcendental subject, we nevertheless saw in chapter 3 that any subject of experience must

also be a free agent, on pain of failing to direct the time-order in which he perceives the Objects of his experiences (on pain of being a mere spectator-agent). Hence a self which, as it were, started life as a subject would have what it takes to acquire free agency in addition. We can therefore speak of the unified self: transcendental in virtue of its agency, empirical in virtue of its subjecthood. This diversity of commitment will be explored further in this chapter. We shall be able to call this self the unified transcendental self, because the transcendental commitments of the self trump its empirical commitments in ways which will become clearer in this and the two following sections.

The transcendental nature of the self is confirmed by the phenomenologies of experience and agency. That the phenomenology of agency reveals a transcendental self is rather obvious. Introspection conducted with the aim of finding out who or what is performing an action alights on nothing in any way empirically characterizable. An agent can certainly introspect his own motives or reasons for action, but those particular motives are not the agents; nor are they essential to the constitution of him as a person: total character change does not in itself affect identity. So long as he thinks of his actions as free, he cannot think of himself as constituted by his reasons for action; he must think of himself (=the owner of his character, but not essentially connected with it) as the agent. The phenomenological point here is that introspection does not reveal one's character as the agent, but something standing behind that character, and free of its trammels. Once his motives and reasons have been subtracted from the

constitution of the agent, there is nothing empirical left for the agent to be characterized by. (I touch again on this issue in 5.5.)

The case of experience requires somewhat closer examination. The general claim I advance is that there is no empirical content either to the subject of representational experience, or to the subject (which is also the object) of non-representational experience. This might seem rather surprising in view of the empirical constitution of the subject; nevertheless it seems to me to be the case, and it is a striking aspect of what I referred to above as the trumping of the empirical by the transcendental (given that a subject of experience is necessarily also a transcendental agent). The subject of experience is a bare point of view - literally so in the case of representational experience (cf. TLP 5.631 - 5.641); metaphorically so in the case of non-representational experience. That the subject is bare is shown in phenomenology by the fact that, as Hume found to his cost, it cannot become an object of representational experience. It can, in the other hand, become an object of non-representational experience, but in so doing it remains bare. The myself whom I feel when I feel a pain in my foot is as devoid of empirical content as the I who feels the pain; or the I who sees a tree.

We should consequently say that the general form of representational experience is:

x perceives: $Fy, -(x=y)$

and of non-representational experience is:

x perceives: $x Mly.$

The variable 'x' here ranges over transcendental subjects, 'y' over

empirical objects. The characterization of representational experience is meant to bring out the fact that the subject of the experience, x , is transcendental (contentless), and the object composed of a particular, y , characterized empirically: $F(\)$. The characterization of non-representational experience is meant to bring out the fact that the subject of the experience, x , is transcendental (contentless), and occurs as object of the experience also. The experience is not merely comprised of a subject's self-awareness, however, but has some empirical content. This content can be formally expressed adverbially, as qualifying the way in which the subject perceives himself: he does so, for example, painfully. The difference between the two regimentations reflects the fact that the way in which the empirical content of non-representational experience attaches to the object (=subject) of that experience is different from the way in which an object of representational experience is characterized by its empirical properties. An object of representational experience is not a transcendental object: all its properties - and especially all its essential properties - are empirical; a transcendental self has no essential empirical properties. (I return to this in 5.3 and 5.4.) The formal regimentation expresses the fact that the subject of representational experience cannot perceive himself, whereas the subject of non-representational experience must do so; it also symbolizes the transcendental status of the subject of representational experience and the subject (=object) of non-representational experience, and the empirical status of the objects of representational experience.

It might be objected at this point that the formal characterization of non-representational experience is regressive. Mackie argues¹ that if a cognition were wholly or partly an awareness of itself, it would have to be an infinite series of cognitions. An awareness of cold would have to be: an awareness of (cold and an awareness of (cold and an awareness of (...), and so on to infinity. Certainly it would be implausible to ascribe to a subject an infinity of awarenesses, if each awareness were logically distinct from its fellow members in the series. But is that the case? Someone who thinks that an awareness of cold is wholly or partly an awareness of itself is hardly going to allow that any member of the series of awarenesses is logically distinct from any other member of it, or indeed logically distinct from from any segment of the series, or from the series as a whole. And there seems no reason why he should: he can claim that a subject possesses any one member of the series iff he possesses the whole series.

Of course Mackie is right to point out that an awareness of cold cannot simply be tantamount to an awareness of (an awareness of (an awareness of (...), since such a characterization fails to incorporate the coldness of the experience, and hence fails to distinguish an experience of cold from an experience of pain. So we must say - as indeed my formalization makes perspicuous - that an awareness of cold is only partly an awareness of itself. But what is the objection to that? That the experiencer is lumbered with an infinity of logically indistinguishable beliefs is no objection, since it is quite unclear how the possession of an infinity of logically indistinguishable beliefs should, in this case, differ

from the possession of a single belief. (There is no possibility here - as there is, for example, in the case of mathematical beliefs which are logically indistinguishable - of an epistemic gap opening between one member of the series and another.) Given this, it seems more natural to articulate an awareness of cold not as a series of awarenesses, as Mackie does, but as a partially reflexive single awareness, i.e.

→ An awareness of cold and an awareness of

The recurrence of subject as object in the formal regimentation need not, then, be regarded as regressive. I do not think, therefore, that Mackie has introduced a valid objection to the sort of analysis of non-representational experience which I have proposed. That analysis legitimately brings out the essentially reflexive nature of such experience. (Note that the reflexive pronoun does not always convey reflexiveness of consciousness: I can see myself in a mirror without being aware that I am seeing myself, i.e. the body of me the seer.²)

It should be observed that the phenomenological point I am making depends on internalism - the doctrine that the subject has authority over the contents of (in this case) his own phenomenological states. This was entirely to be expected. The derivation of the transcendental self presupposed internalism; it is therefore quite unsurprising that phenomenological considerations, in supporting that deduction, should draw on the same presupposition. That they do so is clear from the fact that the bareness of the

experiencing and (reflexively) experienced self only reveals itself to an inward gaze by the subject which is such that the internalist's account of it is correct. The external view has no reason to suppose that the self is other than constituted by the empirical attitudes which it discerns, just as our view of a stone supplies us with no reason to postulate the existence of a transcendental stone: there would be no warrant for an internalist view of stones (see below 5.3). Hence the externalist has no reason to suppose that the reflexively experienced self is not, like the experiencing self, empirically constituted and so precisely not bare. Only the self as the internalist conceives him (as having authority over the contents of his phenomenological states) can certify that the experiencing and (reflexively) experienced self is bare. So the phenomenological claim rests on internalism.

In effect we have been treading a path very similar to that taken by Descartes in his famous cogito. The insight he achieved was that I, the thinker, can deduce my own existence - that is, the existence of myself the thinker - from the sheer thought-episode itself, internalistically certified as present. The argument is stated in terms of the verb cogitare because Descartes found that whatever mental activity he tried to run it on (e.g. my perception of a tree), the only object whose existence he could certainly deduce from the inside was a thinker: in the above example, I can deduce the existence of myself as one who thinks he perceives a tree (not as one who perceives a tree, since that perception may in fact be unavailable to me). In other words, the occurrence of cogitare in the argument is designed to supply Descartes with an indubitable

premiss. What is absolutely certain for the individual (given the internalism which I have suggested is fundamental to our conception not only of consciousness but also of time) is that there is some mentation going on. The reflexive turn of the argument then deduces the existence of the subject of the mentation. And since thinking is the only thing which I cannot 'think away' without thinking myself out of existence, it follows that my essence is to think: *sum res cogitans*. Nothing empirical is needed to constitute the thought-episode which is the basis of the cogito: the self which is encountered in it is empirically bare.

Malcolm has pointed out³ that it is in general invalid to move from the epistemic premisses that I cannot doubt the existence of X and can doubt the existence of Y to the metaphysical conclusion that necessarily: $\neg(X=Y)$. But although it is true that this argument-form can be used to 'prove' invalid sequents, in the present case the inference is valid. That is because it is a consequence of the internalism underlying Descartes' position here that the metaphysical issue of whether or not my existence is indubitable just consists in the epistemic issue of whether or not I can doubt it; and whether or not it is indubitable that I have a body again consists in whether or not I can doubt it. I cannot doubt my own existence as thinker of this thought; so it is indubitable (by anyone) that I exist as thinker of this thought. I can doubt my possession of a body; so it may legitimately be doubted (by anyone) whether I possess a body (as opposed to: merely seem to possess one). Since no one is, or could be, in a position to allay this legitimate doubt (because if anyone were in such a position I would

be, and I am not), it follows that it is possible that I do not have a body. So (one way of putting it) there is at least one possible world containing only my mentation and its subject (myself). That I mentate is therefore the only thing which is essential to myself as subject. Given the internalism expressed by the claim that I have authority over the questions whether or not I am thinking, and whether or not I exist, it follows quite validly that my essence is thinking (understood as including such states as its seeming to me that..., or my deciding to...). The move from the epistemic to the metaphysical is perfectly in order here, because in this case the metaphysical possibilities are constituted by the the epistemic possibilities confronting me.

It is essential to the argument that it be possible to run it in the first-person. I am aware, from the inside, of my mentations as my mentations: that is a fundamental feature of them. To suggest, in the spirit of Lichtenberg's well-known objection, that Descartes is entitled only to a weaker, impersonal formulation (there is thinking going on), and not to the stronger personal formulation (I think) is to ignore two essential properties of mental states: they are had by subjects; and the subjecthood of mental states is both self-announcing and self-certifying. This encapsulates the internalism which is integral to the cogito. The reply to Lichtenberg is not, as has been suggested,⁴ that the impersonal formulation is equally problematic as (i.e. that it is no more available than) the first-personal formulation, but that the first-personal version is made available by the fact that the cogito presupposes or expresses an internalist view; and that view must, in

the first instance, be formulated in the first person.

Third-personal formulations trivialize the cogito and falsify internalism. It is trivial (because circular) to say that a subject can deduce his own existence from the existence of a thought of his. Obviously he can only do so if he exists. This formulation does not capture the insight that I can now deduce my own existence, on no assumptions, from this thought-episode which I am now enjoying. I can arrive at my own existence with certainty and a priori; nothing in the externalist's stock can yield this. The internalist claim that I can be sure of my own existence now becomes false (as I noted in chapter 1) if it is stripped of its indexicality: if a name and a date are substituted for the occurrences of 'I', 'my' and 'now'. Internalism is a transcendentalist position.

Descartes was primarily interested in the epistemological opportunity provided by the capacity of the thinker to think about himself; I am interested here in the metaphysical question: what sort of self does the self alight on in an act of introspection or self-reflection, for example in a non-representational experience? From the phenomenology, the answer must be that the self encountered is empirically bare, or transcendental. Once we have assured ourselves that there is nothing illegitimate in the notion of reflexive experience, and once we have subtracted the empirical content of the non-representational experience (e.g. the cold), what is left in the position of object of experience but a contentless self?

5.2 The Transcendental and Empirical Selves.

Given the existence of the transcendental self, what is its connection with the empirical self? This question has been of great importance in the history of philosophy, at least since Descartes' discovery of the Real Distinction between mind and body, and of the cogito. But it is curious that the doctrine known by the name of 'Cartesian Dualism', and normally taken to incorporate the view that transcendental and empirical selves are completely separate, is not the position espoused by the historical Descartes. On the contrary, Descartes insisted that the soul is not lodged in the body as a pilot in a vessel, that is, as an entirely separate item with no significant connections to the body. Although souls can exist without bodies and vice versa, the position seems to be that there are overall constitutive connections between the soul and the body: they are, as one might put it, constitutively designed for one another. This comes out in the fact that the cogito will run on all sorts of mental act, including those which are body-involving, such as 'I seem to feel a pen in my hand', or 'I feel a pain in (as it seems to me) my leg'. From all such mentations the existence of the subject can be deduced by the subject himself, and the res cogitans which the deduction arrives at is essentially such as to seem to feel a pen in his hand (and so essentially such as to have a body with a hand), to feel a pain in (as it seems) his leg (and so to be capable of feeling pain embodied just there), and so on.

The mental substance which the cogito arrives at has slots for the corporeal: that is the farthest we can go. In any given mental substance the corporeal does not have actually to be slotted in (unembodied existence is possible). That yields a sense in which the mental is more fundamental than the corporeal, but it does not yield the most radical dualism conceivable: complete separateness, i.e. not even constitutive connectedness, of soul and body. Descartes clearly had some degree of difficulty expressing just the sorts of connectedness and separateness he wanted. In two of his letters⁵ he made the attempt, not altogether successfully. He suggests that we should regard the union of soul and body as 'inessential in a sense, not absolutely inessential' (K p 122). Considered as composed of soul and body, the union is an ens per accidens 'because each could subsist apart' (K p 130); but considered as a human being, an ensouled body or embodied soul is an ens per se (ibid). This way of putting it is in danger of looking like a verbal point ('considered as...'), which it is in no way meant to be. In one place, Descartes expresses admirably one of the two constitutive connections: that a human body is fitted up to receive a soul:

It may be objected [i.e it may be correctly objected] that for a human body to be joined to a soul, is not inessential but its very nature; because if a body has all the dispositions required to receive a soul, which it must have if it is to be a strictly human body, then short of a miracle it must be united to a soul. (K p 122)

It is this union which is 'in a sense' inessential, but not 'absolutely' inessential. Further, if a body is fitted up to receive a soul, it is also the case that a soul is fitted up to be received by a body. This is shown not only by the fact, noted above, that the

cogito will run on mentations which are body-involving, but also by the fact, noted at the beginning of 5.1, that the soul arrived at in our deduction is a transcendental agent, and so requires some body through which to exert its will, and some space in which to act; that space will necessarily house particulars, and hence will have all that is required to constitute our agent as a subject of representational experience as well. And the agent's possession of a body will also provide what is necessary to constitute the subject as a subject of non-representational experience. Such a soul does not depend on embodiment for its existence, but if it is not embodied it cannot perform the functions which constitute it as a soul.

We might help out the characterization Descartes is striving for by comparing what David Wiggins has urged with respect to the connection between laughter and the funny.⁶ Although in any particular case laughter is neither necessary nor sufficient for the object of the laughter to be funny, nevertheless there is an overall constitutive connection between them. They are made for one another in this sense: the funny is constitutively such as to merit laughter;⁷ and laughter is constitutively the appropriate response to precisely the funny:

There is no object-independent and property-independent, 'purely phenomenological' or 'purely introspective' account of amusement. And equally there is no saying exactly what the funny is without reference to laughter or amusement or kindred reactions' (ibid)

We might say, in the spirit of Descartes, that laughter and the funny are 'in a sense' inessential to one another, the sense being

this: there can be a genuinely amusing joke which no one (ever) laughs at, and conversely people can laugh at something which is not funny.⁸ But still laughter and the funny are not 'absolutely' inessential to one another, in the sense made clear by Wiggins.

If something like this is the right position to take up on the connection between the transcendental self (the soul) and the empirical self, we emerge with a dualism 'in a sense' but not an 'absolute' dualism. The position in fact ends up being very close to that propounded by Strawson in chapter 3 of his Individuals, with two slight but significant alterations. Strawson argued that the category of person is primitive. A person is constitutively such as to possess mental and physical characteristics. This partially recapitulates the position I have been aiming to describe. Now Strawson allows that a disembodied soul might enjoy a 'logically secondary existence' (p.115); and he appends a brief description of what that existence would involve: along with 'visual and auditory experiences' he tentatively includes 'some quasi-tactual and organic sensations as at present'. This cannot be quite right, since such sensations are body-involving. The distinction between those sensations which are body-involving and those which are not is in fact just the distinction between non-representational experience and representational experience; a disembodied spirit could enjoy only representational experience; such experience would be had from a (possibly moving) location in public space, but without possessing non-representational experience a soul would never need or be entitled to identify any region of that space as occupied by his body. (Conversely, the possession of non-representational experience

would both entitle and compel a soul to think of itself as possessing a body.)

Apart from this adjustment to Strawson's account there is one other that we must make. Strawson suggests that not only should we regard the existence of a disembodied person as logically secondary (i.e. such a person has, as it were, slots for a body) but further that only a sometime embodied person can qualify (later) as a disembodied person. Such a one will live off his memories, or live vicariously by taking a spectatorial interest in human transactions; but

in proportion as memories fade, and this vicarious living palls, to that degree his concept of himself becomes attenuated. At the limit of attenuation there is, from the point of view of his survival as an individual, no difference between the continuance of experience and its cessation.
(p.116)

There are two issues here which are confused. Certainly the possibility of vicarious living must remain with a disembodied person, as it must with an embodied person. For a person to cease to be capable of entertaining (i.e. to cease to be in any sort of position to entertain) the possibility of seeing things from another point of view is just for that person to die. There is no wedge to be inserted between the availability of first- and third-person perspectives to a subject. So if Strawson means to say that as the possibility of vicarious living evanesces (rather than merely becomes boring), the person ceases to exist, his claim is quite correct.

What is not correct is to count in the attenuation of memory as a factor here. There is no reason why a person should not lose all

memory of his embodied state, just as an old man may lose all memory of his childhood, or (an extreme case) may suffer an amnesia severe enough to deprive him of all recollection of himself before a certain date; as in such cases, a disembodied person can be mnemically cut off from his former embodied life without prejudice to his continuing existence as the same person. Consequently, there seems to be no bar to tolerating the possibility of the existence of unembodied persons who are not also disembodied: persons who have never had (and indeed never will have) bodies. Such a possibility does not damage the overall connection between soul and body; as a bare possibility it is just as inoffensive - and for the same reason - as the possibility of a joke which (although there to be appreciated) no one ever appreciates. If we make this adjustment to Strawson's position - and arguably nothing of importance is thereby compromised - we emerge with a view which, contrary to what Strawson himself supposes, is recognisably Cartesian, using this epithet with more of an eye to what Descartes actually thought than is usual nowadays.

In terms of the distinction between the transcendental and empirical self, we might say that within the 'absolutely' unified category of person we can discern transcendental and empirical components. The empirical component comprises such things as a person's desires, beliefs and other 'psychological attitudes (what interlocks with action-explanation), and the conceptual content of his non-representational experiences (the pain in his pain); the transcendental component comprises the subject of his representational experiences, the subject and object of his

non-representational experiences, and the agent of his actions. It is somewhat misleading to call the empirical component a self, since it lacks either subjecthood or agency and so cannot exist on its own (there could not be a particular unowned set of beliefs, desires, pains etc); it is appropriate, on the other hand, to speak of a transcendental self, since unembodied existence for a pure subject, or agent, or both, does seem to be a bare logical possibility: in that sense Descartes was right to accord a certain metaphysical priority to the item arrived at in the cogito. This is the second respect in which the transcendental component in the self trumps the empirical; in spite of the overall constitutive connections between them, the actual existence of a transcendental self does not require the actual existence of an empirical self. But the reverse is not true. The transcendental and the empirical necessarily go together; but the empirical needs the transcendental more than the transcendental needs the empirical.

5.3 The Identity of the Transcendental Self

In Chapter 4 we saw that as far as spatial considerations go, selves and spatiotemporally located objects are on a metaphysical par: both sorts of object come into the category of particulars, and both are capable of constituting particularity insofar as they either are empirical (Objects) or are capable of being embodied (selves). As far as particularity is concerned, a stone and an embodied self are equally part of the empirical order. But particularity is only one consideration. A stone does not have free will. I cannot constitute my own transcendental status without constituting the transcendental status, via the freedom, of any other agent, but that constitution obviously stops short of establishing the transcendental status of particulars which are not agents.

Hence we should talk of the substance of a stone, and not about a transcendental stone underlying the empirical stone. There is no transcendental object because there is nothing more to a stone than an instantiation of a collection of empirical properties at a particular time or times and place or places. If there were a transcendental stone, it would make sense to ask after its identity-conditions and its transcendental properties; and neither of these enquiries does appear to make sense. As for identity-conditions, the stone persists just as long as the collection of its empirical properties (or some reasonable and

perhaps only roughly specifiable subset of them) persists, and no longer. Again, if there were a transcendental stone, it would be intelligible to suppose it capable of 'hooking on' to different instantiations of empirical properties; but this is not intelligible: different instantiations of different sets of stone-properties are just different stones. As regards transcendental properties, it is very hard to see what these could be; the properties inhering in a substance are just its empirical properties. There are no propertyless substrates. To put it in the terms I have been using, we might say that externalism about stones must be correct. We should reject Kantian transcendental idealism, with its noumenal realm of unknowable objects and their properties.

But although we should reject transcendental idealism in its Kantian form, we should not reject every form of the doctrine. For there are transcendental objects: selves are partly composed of such objects. A person is comprised of a transcendental and an empirical component, and in this case it is right to conceive of the connection as the hooking of empirical properties onto a transcendental object, in the service of internalism, with its own identity-conditions and transcendental properties; its transcendental properties are freedom and authority over the contents of its conscious states, including decisions. What account should we give of its identity-conditions? Here we have to consider two questions: what constitutes the identity of a self over time, and what constitutes its unity at an instant?

The identity of the transcendental self over time is constituted in the first instance by the availability of connections

of memory and intention, and its identity at an instant is constituted by the unity of its consciousness. Of course there is no prospect of elimination of either of the terms in these constitutions. Memory presupposes identity as identity presupposes (the possibility of) memory: there is nothing problematic in this mutual presupposition.⁹ Likewise for intention. These two psychological attitudes are the crucial ones for identity: they are the only ones which can connect a self with itself across temporal gaps, and they do so absolutely (e.g. if I remember doing or perceiving something, I remember myself doing or perceiving it). Links of memory and intention are sufficient for personal identity, and the possibility of them is necessary.¹⁰ No other psychological attitudes enjoy this status; they may change as much as you please over time without prejudice to a person's persisting identity. It is conceivable that there should be total interruption or wholesale change of a person's empirical properties, while his continuing identity remains unaffected. It makes sense for someone to have an intention about what he will do in several years' time, even if he knows he will have undergone wholesale empirical change by then, and that the only thing which will survive is the memory of this intention.¹¹ The unity of a consciousness which constitutes a self at an instant is just the capacity of that self to self-ascribe its experiences and thoughts at that instant. Again, there is mutual presupposition between the analysandum and the analysans here.

To make these remarks is simply to draw certain analytical connections. We have said nothing which an externalist need deny: for we have not yet said what it is to have a memory or an

intention, or what the ability of the self to self-ascribe its experiences consists in. But since we are enquiring here into the commitments of the notion of the transcendental self, and since the derivation of that self presupposed internalism, it is the internalist reading of the analytical connections which interests us here. (In the case of memory, I am doubly committed to an internalist construal, because the support I gave to the internalist view of perception in chapter 1 must carry straight over to the case of memory, memory being a special case of - or at least depending on the use of - perception.) Now the internalist will obviously deny that the criteria for identity of selves over time and at an instant consist in criteria for bodily identity: on his view any transcendental self, although necessarily embodiabile (5.2), is not identical with any particular body it enters. If the internalist is right, we must at least be able to make sense of the possibility that the rule 'one person-one body, and vice versa' should break down over time and at an instant. We must be able to envisage the individuation of transcendental selves cutting across the individuation of bodies. Now we are ordinarily extremely reluctant to allow exceptions to the rule, because we suppose that bodily identity and personal identity at least supervene on one another; but, as I have stressed, supervenience is a relation we do not understand, and perhaps because of that (or, as I should prefer to say, it is a constraint on our understanding of supervenience that) we can conceive of particular superveniences coming adrift. As elsewhere, supervenience serves here as a marker for commitments which we do not know how to reconcile, but which we find it hard to

relinquish.¹² Supervenience uneasily splits the difference between an externalist identity theory, and an internalist dualism.

It has been thought that there is no chance of setting up coherent criteria of identity for transcendental selves which do not reduce to criteria of identity for particular bodies. Taking identity over time first, it was argued famously by Kant¹³ that 'I' cannot refer to a soul-substance, since as far as that goes there could be a succession of such substances over time - each passing its consciousness (including memories and intentions) to the next in the sequence - and no one would be any the wiser. In that situation, 'I' would not refer to an individual soul-substance in the sequence, but would have to refer (if at all, which Kant actually denies) to the sequence as a whole. Strawson suggests that the incoherence lies in the failure of the Cartesian (the internalist) to explain what entitles him to believe that he is just one perduring soul-substance, rather than a succession of soul-substances passing consciousness on like a baton in a relay-race. What would distinguish these scenarios? The claim is that the Cartesian will want and have to distinguish these scenarios - but spuriously, since there is no distinction between them.

How finely are psychological attitudes being cut in this challenge? If numerically the same memories and intentions get transferred from one soul-substance to the next, then while it is true that there is no genuine distinction between the two scenarios, it is not true that the Cartesian has to distinguish them. In speaking of the transfer of consciousness from one fleeting soul-substance to another, and in smuggling persisting memory and

intention into the constituents of such a travelling consciousness, one is simply redescribing the situation in which a single soul-substance persists over time. For memory and intention to persist is for the self to persist; if memories and intentions are transferred, everything sufficient for personal identity over time is transferred with them. It is an agreed datum between the Cartesian and the objector that the hypothesis of separate, fleeting soul-substances, as described (i.e. with persistence of memory and intention), is empty. The Cartesian is quite entitled to fail, therefore, to distinguish this scenario from that of the perduring self; for they are the same scenario.

But surely what the Kantian challenger has in mind is that the internalist ought to make room for the possibility that memories and intentions which I take to persist do not really do so, but are continuously perishing, while qualitatively identical but numerically distinct memories and intentions are continuously born to carry the baton forward. Now the internalist must indeed countenance this possibility, but I do not think he is thereby led into making spurious distinctions. Consider first that either the memories which go to make up the envisaged succession are accidentally concatenated or they are not. If they are, then it will be possible to interpose between one set of psychological attitudes in the succession and the next, so that the later set will really be mnemically cut off from the earlier; and in that case it will be coherent (although not mandatory) to say that we have here a succession of distinct selves, rather than one persisting self. But if there are non-accidental links between one psychological set and

the 'next' (whatever those links are - perhaps causal), such that each link in the succession is non-accidentally responsible for the tranference of its qualitative characteristics to the 'next' link, then the right view, on anyone's showing, will be that we have one persisting self after all. For, one might say, to have a succession of qualitatively identical memories and intentions suitably and non-accidentally linked just is for numerically identical memories and intentions - and therefore for the self - to persist.

Now to distinguish these two possibilities is not to avert the Kantian challenge, which can be restated as follows. How can a subject rebut the sceptical doubt that successive psychological sets are not in fact appropriately linked, so that they are genuinely numerically distinct? And then how can he rebut the consequential sceptical possibility that at each stage a new self is born with its own proprietary set of psychological attitudes numerically distinct from, but qualitatively identical to, the previous set? Here we touch the core of the Kantian attack on the Cartesian's conception of the soul. But although internalism certainly comes under intense pressure at this point, it is not clear that it buckles. On the internalist view, I have to suppose that the connection between myself and the present moment is so close that I cannot allay the doubt that perhaps I exist in this moment and at no other time. That the free transcendental self should thus stand under the tyranny of the present moment is perhaps to be expected, after our deduction of the former on the basis of the latter. The free self which constitutes the present moment as present finds itself trapped, in a sense, in that moment.

But only in a sense. The scepticism to which the subject is liable is of a piece with other epistemological scepticisms which, as we have seen, cannot be rebutted but cannot, as such, vitiate our right to operate with the concepts whose application they call in question. (As I put it earlier, epistemological scepticism cannot be upgraded to metaphysical scepticism.) The subject has, on the internalist view, authority over the contents of his intentions and of his apparent memories; given this authority, he can at least be confident that he has intentions and apparent memories, that is, mental states enjoying constitutive connections with the future and the past respectively. So just as the present moment enjoys constitutive connections with the future and the past, even if no future and past in point of fact exist, so a free transcendental self is essentially such as to entertain intentions concerning the future and have memories of its past, even if it actually exists for no more than the present moment. In constituting the present moment, the free self constituted the A-series as a whole.

The subject has no criterion for settling which of his apparent memories are veridical (although considerations of coherence will exert a negative influence, since coherence is a necessary, but not sufficient, condition for veridicality). Because memory is a species of perception, neither the subject nor anyone else can use causal considerations to determine which of his apparent memories are veridical. Earlier, I offered an account of veridicality of perception in terms of indefinite corroborability of experiences, but in the case of memory this analytical connection is of less practical use, since the past cannot be revisited. Of course the

corroboration of perception depends on the use of memory; so there is less of a gap between the epistemological statuses of perception and memory than might appear.

Ordinarily I take it that my apparent memories are veridical - that I have retained my original perceptions (to put it in terms of the recursive model of memory) - and, consequently, that I have persisted during the period from original perception to present memory. Similarly, I assume that when tomorrow I shall enjoy apparent memories of what I did today, those apparent memories will be veridical, and that the subject who enjoys them will accordingly be me. There is no dislodging the epistemological sceptic who queries my persistence, and the veridicality of my apparent memories; but, equally, such a sceptic cannot upset my right to assume that I do really persist and genuinely remember: that 'I' on my lips refers to the same self yesterday as it does today and will tomorrow, and that my apparent memories from day to day include veridical memories which therefore retain numerically, and not merely qualitatively, the same perceptions as the originals (and so are numerically the same memories persisting over time). Once again, the striking fact is not that the sceptic might be right, but that he might be wrong - not that I might be a fleeting soul, non-existent until now and about to pass out of existence again, but that I might not be so trammelled; that what I am in business to do, as it were, whether I succeed or not, is to endure.

I noted above that if I suppose myself to be situated at one stage in a succession of numerically distinct psychological sets, it is coherent but not mandatory to conceive of myself as about to

perish, and that a distinct self will take over, with a new psychological set. The scepticism explored above turned on this possibility. That conception is not mandatory, because it is equally coherent to conceive of myself as persisting through complete psychological change, including breaks of memory and intention. Whereupon the question naturally arises: how am I to be able to tell the difference between the two types of scenario? And, indeed, how am I to distinguish either from the ordinarily assumed situation in which I persist with numerically the same memories and intentions? It seems clear that the internalist can provide no answer to these questions. On his view the persistence of the transcendental self does not consist in anything non-transcendental: there could be no empirical evidence or criterion for or against it in any particular case. It is simply barely and transcendently conceivable that just one of the above three possibilities should obtain. The empirical facts here radically underdetermine the transcendental facts. That, again, was to be expected: the empirical temporal facts do not determine which moment is now; likewise the empirical personal facts do not determine who I am over time.

What about the identity of the self at an instant? Strawson attacks the Cartesian account on the grounds that it should be able, absurdly, to distinguish the scenario in which just one Cartesian self thinks and speaks, from that in which a thousand souls think and speak in unison.¹⁴ The form of this argument ('The Argument From A Thousand Souls') is similar to that of the previous argument ('The Baton Argument'). And since the Argument From A Thousand Souls has the same overall form, the same riposte works against it. If each of

the thousand souls is a unified consciousness psychologically cut off from (i.e. with no connections of memory and intention to) each of the other souls, then indeed the scenario in which a thousand souls speak through my mouth is genuinely different from that in which just one soul speaks through my mouth. Once again we need an internalist perspective to yield a distinction between the situation in which each of the thousand souls is a distinct soul, and that in which what we have is simply different aspects of a single soul, psychologically cut off from one another.

To envisage a thousand souls sharing my body is just to push the possibility of schizophrenia to an extreme. To envisage two souls sharing my body is to envisage the possibility of one soul having distinct memories and intentions from the other, and that only one of the souls is me. Supposing the souls occupy the same body on alternate days, it is to envisage the possibility (for example) of one of the souls (me, say) declaring on an odd-numbered day 'I will always brush my teeth before going to bed', and then proceeding to remember, and act on, this intention on odd-numbered days. On even-numbered days, on the other hand, the soul occupying the body who is not me remembers no such intention. What makes it the case here that we have two souls occupying one body, as opposed to one soul with the mnemonic peculiarity that that on even-numbered days he can only remember what he did on other even-numbered days, and similarly, mutatis mutandis, for odd-numbered days? Well, just that and nothing else; only an internalist view can make sense of the difference between the scenarios. I can imagine its being me all along, or its being me only on odd-numbered days. That difference

cannot be supposed to consist in anything: the empirical facts again underdetermine the transcendental facts.

But if the empirical underdetermines the transcendental, it does not follow that the transcendental can simply cast itself entirely adrift from the empirical. This is clear from the identity-conditions of the transcendental self, i.e. from what, in general and analytic terms, the difference between one transcendental self and two is: in the domain of agency, the distinction is supplied by the possibility of one transcendental self performing (and intending to perform) acts not performed (or intended) by the would-be distinct self; in the domain of experience, the distinction is supplied by the possibility of the transcendental self perceiving (and remembering that he has perceived) a sector of the world unperceived by the would-be distinct self. That is the principle of distinction between one transcendental self and two; and it is clear from its statement that two selves cannot be distinguished at the transcendental level without being capable of being distinct, although as we have seen not necessarily capable of being distinguished, at the empirical level. What this shows is that while the transcendental self is not individuated by empirical criteria, still the possibility of an empirical difference between two transcendental selves is essential to their transcendental distinctness. The transcendental is vitally, but still indirectly, connected to the empirical.

Recall that in chapter 4 we concluded that space is an empirical phenomenon, because the particularity of a spatially located object is ultimately guaranteed by the possibility of

something empirical distinguishing it from any purported duplicate. We said that a universe does not actually have to be empirically distinguishable from a duplicate: in that sense the empirical facts can underdetermine the distinction between a universe and its duplicate. We did not want to express this underdetermination by saying that the distinction between a universe and its duplicate is transcendental, because there is no guarantee that the empirical facts will underdetermine the distinction. But there is a guarantee that they will underdetermine which moment is the present moment; we can express this difference by saying that space is empirical and time is transcendental. But now from our discussion of the transcendental self, it is clear that in terms of its identity-conditions, 'I' is on the side of 'here' rather than 'now'; we have arrived at an essential meeting point between the empirical and the transcendental. The principle of distinctness between two transcendental selves - that one should be able to perceive something different, will differently, and have different memories and perceptions from the other - amounts to the demand that the distinctness at least be capable of making an empirical impact somewhere. Of course the difference between one transcendental self and two need not be empirically detectable, but it must be empirically realisable. That is just what we should say too about the difference between two places. And then the unsullied transcendental status of the present moment comes out in the fact that it need not - indeed cannot - make any empirical impact anywhere.

We should accordingly be careful in how we express the way in

which the empirical affects the transcendental. Although the question of the identity of the transcendental self cannot cut loose from empirical considerations, this is a far cry from the proposition that its identity is even capable of being settled empirically, still less that it is reducible to, or consists in, the identity of something empirical, such as a body. The general and analytical connection between the transcendental and the empirical does not, and cannot, require empirically detectable repercussions of transcendental differences, let alone specific and direct connections between particular transcendental selves and particular bodies.¹⁵ In practice we are reluctant to employ separate criteria of identity for selves and bodies. But that is because we are pulled in an externalist direction as much as an internalist: we feel that selves ought really to be individuated by bodies, but we cannot (at least, I cannot) fault the derivation which yields an internalist view of the self, the consequences of which are that although what happens to me must be capable of making an empirical impact, what happens at the empirical level has as such nothing to do with what happens to me.

The fact that the transcendental has commitments to the empirical while retaining its own integrity is strongly reminiscent of what we found we had to say about freedom in chapter 3. We have seen that in respect of phenomenology (5.1), existence (5.2), and now identity-conditions the transcendental component in the self trumps the empirical component, in the sense that transcendental differences can fail to register at the level of the empirically detectable; but we have also seen that the transcendental self

retains overall connections with the empirical; and that what internalism forces us to say about the self has several elements which run counter to externalist intuitions which we undeniably have. I should say that this tension within our concept of the self was to be expected after the discovery of contradictions lurking within the concept of freedom: freedom requires rationality, but rationality undermines freedom. The transcendental self was arrived at in the deduction of freedom; so it is unsurprising that it inherits from its governing concept of freedom something of the latter's status as a house divided against itself. Surely analogous difficulties to those consequential upon the freedom of the transcendental self lurk behind the possibility of its embodiment. In ordinary life we work with an externalist concept of personal identity - we ordinarily take it to consist in bodily identity - but we cannot fault the coherence of the internalist's view that they could conceivably come apart. If that is so, it is not clear to me how, or even that, this division of loyalty will be resolvable. But there is an important problem relating to the self which I think its divergent commitments actually help us to solve, and to which I now turn: privacy.

5.4 The Question of Privacy.

If the distinction between the transcendental and the empirical in experience which I have proposed (5.1) is correct, it yields us a short way with the question: is experience private, that is, logically inaccessible to others? The answer must be that the empirical content of my experience is equally available to others as to myself. In the case of representational experience what we must say is that although a particular experience is proprietary to its subject (since it is individuated by its subject), nevertheless the particular empirical content of my experience (e.g. the table) does not incorporate myself as part of it, so that there can be no logical bar to precisely the same content being available to be enjoyed by someone else (cf. 1.4). In the case of non-representational experience, although again the particular experience is proprietary to the subject-object (since, again, that is how experiences are individuated), and although - more seriously - the particular object of such an experience is proprietary to its subject (since it is identical with its subject), nevertheless the general empirical content of my experience (the precise quality of the pain, say) is equally available to others as to myself.

The point is that it is just an easy consequence of the way I have characterized representational and non-representational experience in 5.1 that no problem of privacy arises. Given that the subject of both sorts of experience - and in the case of non-representational experience that means the subject-object - is transcendental; that just releases a purely empirical component in

the experience which is detachable from the experience and is logically available to be experienced by any (transcendental) subject. In the case of representational experience, that component is particular, since the particular object of perception does not incorporate the subject of that perception; in the case of non-representational experience, that component is general, since while the particular experience does incorporate the subject in the object, that object has a general empirical component which is available for public release.

Of course, merely to point out that the favoured way of regimenting experience obviates a problem of 'other minds' is not to justify that regimentation or dispose of the problem. The problem of privacy will then just shift its sights onto the availability of that regimentation itself. I cannot here hope to meet all the objections to my regimentation which might conceivably be raised. I hope that I have already rendered that regimentation independently attractive. I have already countered the claim that the characterization of non-representational experience is regressive, and I have suggested that the characterization receives some support from phenomenological considerations. But I take myself to have provided an argument to the effect that the self which figures as subject of experience is transcendental (4.2, 5.1), and I have rebutted the suggestion that my theory of knowledge entails the privacy of objects of representational experience (1.4), so that any remaining problem will concern the entitlement to detach a transcendental object (=subject) from an empirical object in non-representational experience.

In order to meet this problem, we must remind ourselves what the distinction between the two types of experience is, and assure ourselves that it is non-arbitrary. It will help to concentrate our thoughts if we examine a passage of A.J.Ayer's in which he argues that the distinction between the types of experience is arbitrary. Ayer considers the question of the availability not only of 'public' sensations, but also (the complementary case) of 'private' tables, and argues that our settled habit of describing sensations as private and tables as public is merely a contingent artefact of our notation.

Tables are public; it makes sense to say that several people are perceiving the same table. Headaches are private: it does not make sense to say that several people are feeling the same headache. But just as we can assimilate tables to headaches by introducing a notation in which two different persons' perceiving the same table becomes a matter of their each sensing their own private 'tabular' sense-data, so we could assimilate headaches to tables by introducing a notation in which it was correct to speak of a common headache...A London particular might come to be a local headache as well as, or instead of, a local fog...The conditions which would make this way of speaking useful do not, indeed, obtain; but that they do not is, once again, a purely contingent fact.
([2],pp.49-50.)

But it is surely clear that in Ayer's fantasy case of the London headache (just as in Wittgenstein's fantasy of pain-patches), individuals would still feel the pain in their respective heads, and not in the public places where the fog used to linger. They could certainly conceive of the public places as causally responsible for all their individual headaches, but those places would not figure as intentional objects of their mental states: intentional objects would remain dispersed among the sufferers and so proprietary to each person; it would still be the case that nobody else could feel

my headaches. The best way to represent this inevitable, as it seems to me, view is to incorporate the subject as object of the experience.

It is by no means an arbitrary matter which sorts of experience we categorize as representational and which as non-representational; pain and other experiences which we put into the latter category cannot go public; that is, their objects cannot be made available to be experienced by more than one subject. Conversely, Ayer's privatizing manoeuvre completely fails to change the public status of the objects of representational experience. The so-called private 'tabular' sense-data would not in fact be private in respect of their content. Of course objects of representational experience can, if one so stipulates, be individuated in such a way that no two persons share the same perceived object. But to set the terms up in this way is not only to introduce chaos into one's theory of knowledge, it is also to misrepresent the character of the experience. It is to renege on the fact that table-experience is, precisely, representational: that is to say, the table occurs as object of the experience, but the subject does not. There is nothing in the content of the experience to warrant privatization; and the move is seriously misleading as to the epistemological facts of the matter. In sum, there could be no such thing as the representational experience of a headache, or the non-representational experience of a table.

Given now that there is a non-arbitrary distinction between the two types of experience, and that the distinction is correctly captured by the regimentation proposed earlier (5.1), it follows

that since the subject of experience is transcendental, the object of non-representational experience (which is identical with the subject) must also be transcendental. Hence it is detachable from the general empirical component (the character of the feeling); and hence the possibility of setting up a subject of logically private experience is nullified. Any subject is saddled with his own transcendental status, and hence with the bareness of himself not only as subject but also as object of non-representational experience. He does not figure in the quality of that experience. The empirical component of that experience is now cast adrift to be picked up by other transcendental subjects; for a subject cannot suppose that he uniquely enjoys the status of subject; this is because what constitutes him as a transcendental subject (his agency) is equally available, in principle, to constitute other such subjects.

According to this line, the particular objects of particular veridical visual experiences, for example, are available to anyone, in the sense that had another person been appropriately positioned at the appropriate time, he would have seen precisely the same (and not just precisely similar) objects. Objects of veridical visual experience are waiting around to be picked up, and whoever picks them up can own them, as it were. The ownership of the objects of veridical representational experiences, like the ownership of books, is a purely contingent affair. Ownership of the objects of non-representational experience, on the other hand, is necessary ownership; so that only the general empirical quality, and not the particular objects of my non-representational experiences, could be

experienced by others: no one else could have or have had my pains.¹⁶ To have someone else's pain. i.e. to have an experience of the form:

x feels: in pain(y), $-(x=y)$

is not the form of a possible experience. If I feel anyone's pain, it is necessarily my pain which I feel. That is what is meant by the 'logical' ownership of my own pain.¹⁷

The objects of non-representational experiences are necessarily individuated by subjects; those subjects recur as (partial) objects of their experiences. To the objection that such things as coughs are equally owned by unique subjects and hence that the whole business of ownership is just a matter of grammatical convention, the reply is that on the contrary there is a deep metaphysical difference between ownership of pains and ownership of coughs. It is simply incorrect to say, as Ayer does,¹⁸ that 'if you cannot have my feelings, it is equally true that you cannot sleep my sleep, or smile my smile, or speak with my voice.' You can do all of these things, simply because we do not individuate voices, smiles and sleeps by persons but by bodies. Two people may then share a sleep or a voice by sharing a body (cf 5.3); the same cannot go for pain: a particular centre of consciousness figures ineliminably in the content of non-representational experience. Even where two persons share a single body, or part of a body (Siamese twins), although they may both feel pains at the same time, in the same place, and of identical quality, it will still be right to speak of two numerically distinct pains. But 'logical' ownership does not pose a threat of privacy: although the particular objects of

non-representational experience are proprietary to their subjects, so that no one else can feel my pains, this rather gross fact conceals the diversity within the accusative of non-representational experience. There is a transcendental component; and there is an empirical component which is general - hence detachable and available to be picked up by other subjects of experience. It seems to me that only some such account can deliver the right answer to the question of our justification for belief in other minds.

The traditional Wittgensteinian polemic is in my view quite impotent to dislodge the private linguist from his position, simply because it begs the question against him. It fails to take seriously the possibility of a really thoroughgoing privacy, by assuming that any purported private linguist will have to allow publicity into his language somewhere: his position is then easily and predictably undermined from the very position where he admits publicity. I shall briefly examine this failure with respect to two polemics which it is possible to discern in the pages of the Philosophical Investigations. (Nothing here turns on the distinction between private language and private experience: the one is available if and only if the other is.)

In the first place, there seems to be an argument to the effect that the private linguist is not following rules in his private language, as opposed to appearing to follow rules; or, he may think he is confronting the same sensation again on a later occasion, but he cannot give content to the assertion that he really is confronting the same sensation, as opposed to merely seeming to. He cannot draw the appropriate 'is right/seems right' distinction. Some

of Wittgenstein's remarks seem to suggest that the private linguist cannot draw this distinction because he is liable to a scepticism about memory: if he thinks 'this is sensation S again', he cannot guarantee that his memory of the sensation he called 'S' is not defective. The trouble with this line is that it proves too much. Private language is no more vulnerable to scepticism about memory - or in general about our ways of gaining knowledge of the past - than is public language. As far as that goes, we are all in the position of having only more copies of the same newspaper to go on.

Instead of offering a sceptical point, we might read Wittgenstein as arguing that the private linguist has no reason to suppose that he is following rules in his linguistic practice at all, as opposed to thinking he is. Here is one place where it is common among commentators to beg the question against the private linguist. Wright, for example, suggests that

nothing counts as giving yourself, or anyone else, reason to take certain episodes of your behaviour as genuine cases of rule-following which does not consist in supplying reason for thinking that it could reasonably be so interpreted by others ([2], p.220-1)

But, plainly, the private linguist not only can but must reply to this that insofar as he can supply reasons or justifications for his judgments ('this is 'S' again') these reasons and justifications will perforce be given in a private language too. If judgments about sensations can be private, reasons for taking these judgments to be rule-governed can be private too. And must be: it would be disastrous and incoherent to suppose that private judgments could be underpinned by public justifications.¹⁹

It is possible to detect here and there in Wittgenstein's writings materials for another sort of attack on privacy. This line argues that there are constitutive connections between the inner and the outer. The connections are of two sorts: typical patterns of causation and typical patterns of behaviour. In the case of pain, not any old event can cause pain; and if events of an appropriate sort occur, and given a suitable causal field, pain will be produced. This is, if you like, a synthetic a priori truth.²⁰ Pain is constitutively connected with its typical causation. This point is gestured at by Wittgenstein when he writes:

If I see someone writhing in pain with evident cause I do not think: all the same, his feelings are hidden from me. (PI p.223)

Much fuller treatment is accorded by Wittgenstein to the connection between pain and pain-behaviour. Here again the point seems to be that there is an overall constitutive connection between the two phenomena: screwing up one's face is a piece of behaviour which is fitted up, or made for, the expression of pain. Pain-behaviour is, in some sense, a criterion for pain.²¹

Using these materials, one may construct a position (which may or may not be Wittgensteinian) according to which pain cannot be private because it is constitutively connected to other phenomena - certain kinds of causation and behaviour - which are inevitably public. The point can readily be generalized from the case of pain to other sorts of non-representational experience. In the first place, some of these have their own connections with typical patterns of behaviour (e.g. tickling); but those which do not

nevertheless lie on a continuum, or are located in a 'space', in which pain is also located; so that they cannot avoid having constitutive connections - albeit indirect ones - with the outer. Interestingly enough, the position we are characterizing differs only slightly from Mill's picture of the connection between the inner and the outer. Mill thought that I can observe the tripartite structure of causation, sensation and behaviour working in my own case, but that I need an analogy to apply the whole structure to others, since in their case I can observe only the first and third components (pp.243-4). The objection to this position has generally been that my belief in other minds stands in need on some better grounding than mere analogy; and that no such warrant is in fact available. In the Wittgensteinian application of the structure, no distinction is made between the realization of the structure in my case and that of others. That is indeed the point: the constitutive connections obtain whoever instantiates them. So the need for anything like an analogy apparently does not arise.

Of course the problem which the Argument From Analogy was intended to deal with has not been solved by the Wittgensteinian manoeuvre, but merely shifted. The difficulty now centres on one's entitlement to instantiate the Wittgensteinian version - rather than the Millian - of the tripartite structure. Is there a univocal notion of pain? This problem is less pressing for someone who takes my line on the nature of non-representational experience, even although, paradoxically, it is a corollary of that view that there is a sense in which such experience is private: namely no one else can have the particular objects of my non-representational

experiences (because the subject recurs as object, and no one else can be me). But because, on my line, the subject in question is transcendental, the privacy is not such as to precipitate an 'other minds' problem.

The problem does not just move elsewhere (entitlement to operate with the relevant concept of transcendence), because the transcendental status of the self is independently assured; the basis of the deduction given in chapter 3 does not allow for anything like a 'logical' privacy to get going. It does not allow that the deduced selves might be logically unique: any agent (acting in time) is thereby constituted as a transcendental agent; any subject of experience (capable of acting in the space housing the objects of his experience) is thereby constituted as a transcendental self. Action in time is the deductive base of the derivation of the transcendental self; there is simply no place for uniqueness to intrude here. The deduction does not permit the granting of special status to any of my experiences, because they force me to recognize my own transcendental status, and hence the availability of what is empirical in my experience to be picked up by any other transcendental selves, of whose availability the terms of the deduction also assures me.

It is quite unclear how Wittgenstein could respond to someone who, perhaps out of sympathy for Mill, contested the right to a univocal pain-predicate. Since, as Wittgenstein himself makes terribly clear, the basis of self-ascription of pain is different from the basis of other-ascription of pain, there can be no easy assumption of univocity here. Perhaps the language of pain is keyed

to the outside components of the tripartite structure (causation, behaviour), and not to its middle term (the pain) - this seems to me to be in a way correct²² - so that it is then open to the private linguist, for all Wittgenstein shows to the contrary, to invent a name for the sensation on its own without regard to the public language of pain. Still, it might be said, others could surely know when he was in pain. Suppose then the private linguist privatizes the whole structure: this is another place where the question has often been begged against him, since it has been assumed that he cannot do this. What now is the objection to patterns of causation, pains, and patterns of behaviour which are toto caelo cognitively inaccessible to more than one person? Nothing in the idea of a tripartite structure, and nothing in the nature of the outside components, militates against the privatizing move. Why should not a person react quite incomprehensibly (logically so) to private patterns of causation (in the sense of objects having inscrutable effects on him), and behave in equally inscrutable ways? Only the transcendental status of the self is competent to rebut such a radical and epistemologically catastrophic privacy: it does so while accepting that there is a sense in which non-representational experience is private. By conceding a little, it preserves the citadel.

5.5 Vere's Choice

Russell once remarked that a logical theory may be tested by its capacity for dealing with puzzles; I am going to conclude my account of the self by seeing how it applies to an interesting problem that has been discussed recently by a number of writers,²³ and which has been prompted by reflection on Herman Melville's story Billy Budd. In that story, set in the Napoleonic Wars, Melville describes how Budd, a foretopman of H.M.S. Indomitable, is calumniously accused in the presence of the captain, Edward Vere, of incitement to mutiny by the master-at-arms, John Claggart. Unable to respond because of a speech impediment induced by outrage at this wholly unexpected and baseless charge, Budd strikes Claggart and kills him. Whereupon Vere convenes a drum-head court which, under his suasion and somewhat reluctantly, condemns Budd to immediate execution in accordance with the terms of the recently introduced Mutiny Act.

None of the protagonists of the trial, least of all Vere, is unaware of the deeply regrettable nature of Budd's case; indeed Vere himself feels the intractability of the position acutely, caught as he is in a dilemma between the demands of military duty (the more compelling in the light of the recent mutinies at Spithead and The Nore) and natural justice ('innocent before God' is Vere's own description of Budd). It is clear that Vere could have deflected the court towards something less harsh than sentence of death, had he so resolved, or at least (a less radical possibility favoured privately by the ship's surgeon) he could have placed Budd in solitary confinement, waited until the Indomitable rejoined the main fleet, and then constituted a regular court martial under the presidency of

the Admiral. The story contains the materials for an adverse judgment of Vere, if we wish so to take them, but at the same time Melville warns us against too easy a condemnation of one who had to make a difficult decision quickly, and in the face of an unpredictable and dangerous enemy - an enemy possibly abetted by disaffection among Vere's own sailors, many of whom were impressed men.

Reflecting on the richness of the moral stage-setting of the story (which must of necessity receive scanty treatment here) it is possible to regard Vere as having made the right decision given the situation with which he was confronted, but still to feel that one would oneself have been unable - morally unable - to take a similar decision in similar circumstances. This is the line taken by Peter Winch in his sympathetic discussion of the case, and it has much intuitive appeal. It is a position which, even if it is not everyone's (it is not mine), I think we should nevertheless be prepared to accommodate. The difficulty is that it generates a clash when set alongside something which equally has claims on our assent, namely the principle of universalizability, at least in the version which states that if it is right (wrong) for a man to V in circumstances C, then it must be right (wrong) for anyone to V in all circumstances where the morally relevant considerations are the same. This has the look of an a priori truth about moral reasoning. I want both to endorse the principle of universalizability in this form and to find room for a position such as Winch's. I shall suggest that the apparent incompatibility between them can be resolved only by applying the distinction between the empirical and

transcendental selves.

One way of resolving the puzzle which does not seem to be viable is Winch's own resort of construing the act of decision as itself contributing to the moral character of the action decided upon:

A man in a situation like Vere's has to decide between two courses of action; but he is not merely concerned to do something, but also to find out what is the right thing for him to do...What I have suggested is that the deciding what to do is, in a situation like this, itself a sort of finding out what is the right thing to do; whereas I think a writer like Sidgwick would have to say that the decision is one thing, the finding out quite another. (p.209)

But Sidgwick would be right. Winch's preferred manoeuvre will not work because it does not allow for a certain sort of mistake, namely when one makes the wrong decision having in all good conscience surveyed the moral territory. Winch has to suppose that the only ways of making a mistaken moral decision are failure to appreciate the right moral issues (either through failure to appreciate that the situation was a moral one, or through failure to extract the correct issues from it), or the succumbing to some mauvaise foi. But these do not, I submit, exhaust the field, or even admit the most usual type of case. Why should not Vere - or anyone - come later to suppose that although he was alive to the right moral issues, and lacking in self-delusion, he nevertheless made the wrong decision in condemning Budd? He appreciated the moral issues of the situation but mistook their precise import. (We might be tempted to refuse to allow a gap between seeing the right moral issues of a situation and making the right decision; but that would just be to prescind from the non-cognitivist model Winch is trying to run.) And how could

Vere comfort himself in his subsequent self-criticism with the thought that at any rate that was what he decided, as if just that fact conferred moral respectability on what he did?

It is really another way of expressing this point to say that deciding how to act cannot be a way of finding out how one should act because it would render action itself impossible. It is no answer to a man who does not know what to do to say to him: decide! This is never an answer to the question he is asking; it is at best an injunction to act in spite of the lack of an answer to that question. (It means: don't think, just act!) What the man wanted was to find out what he ought to do, and then act. If the decision were itself a way of finding out what to do he would be in no way helped. To find out, we are told, he must decide. Well, but now which decision is he to make? To find that out he must decide...The only way out of this circle (and it is a circle which confronts the agent, not merely the theorizer) would be simply to short-circuit reflection and act. But the chosen action would not have been decided on in the relevant sense: it would have been chosen arbitrarily - only because it was an action, not because it was that action. There is nothing but confusion to be met with along this path. We should accordingly insist that the cognitive and the practical are, at least in the sense we are here concerned with, toto caelo distinct.

Part of Winch's motivation for taking his line was the consideration that how a situation strikes an agent may not be separable from his inclination to come to a particular decision (p.213-4). Its striking him in that way, one might say, is partially

constituted by his feeling disposed to decide in such-and-such a way. And then surely it is conceivable that situations which are similar in morally relevant ways may strike different people differently, even if they are all responding sympathetically and adequately to the moral realities in front of them. It may simply be 'essentially contestable' what to do in situations of that sort.

Now certainly we can use the phrase 'how the situation strikes the agent' in Winch's way; but if we do so, I think we blur a needed distinction. It is surely possible to express the morally relevant considerations in, for example, the case of Billy Budd, before one raises the matter of how they strike the protagonists. Certainly the morally relevant circumstances over which we universalize must not, on pain of triviality, include the agent's inclination to come to some particular decision. That is a valuable point, and one well taken by Winch. But it does not follow from this that while the morally relevant circumstances are held fixed across two distinct scenarios, say, they can with justification strike the respective agents (or suitably informed spectators) differently. After all, what strikes an agent is surely something in the world, so that if one agent is right to be struck in one way, and another agent right to be struck in another way, that surely suggests that the realities they are confronting are relevantly different, contrary to hypothesis. And if they are not different, surely one agent is missing something, or deceiving himself, or being insincere, or...

Alternatively, if we key the phrase 'how things strike the agent' not indeed to the decision he is disposed to come to, and hence implicitly (I have suggested) to how things are in the moral

realm, but instead to aspects of his emotional make-up and character (including such things as his social position, and expectations entertained by others regarding his behaviour); then I think it is quite unclear that how things strike the agent in that sense should not (if morally relevant) be incorporated into the universalization base. We must be careful here. I am not saying that every aspect (or even most aspects) of Vere's character (say) must be regarded as fixed parameters for the purposes of his decision and universalization. That would drain the life out of Vere's decision (it would stop self-appraisal) and the interest out of universalization. It is quite fair to point out that Vere could not reasonably treat

his own character, and his duty and function as a naval captain holding the King's Commission and his view of that duty and function, as among the realities....For the purposes of the future [i.e. the decision] from then on even these things needed reappraisal...(Wiggins [3], p.180)

But what I am saying is that the morally relevant aspects of the whole situation - including any aspects of Vere's character which are morally relevant - do constitute both Vere's decision base and the universalization base. What the morally relevant features of the situation are is precisely what is up for decision. Vere cannot in advance exclude any aspect of his character and situation from moral reappraisal: Wiggins is right there. But that process of reappraisal may alight on certain aspects of character and dignify them as moral parameters to the debate. If we both insist that features of Vere's character can be morally relevant, and we remind ourselves that only a small fraction of the total number of facts of the situation will

be morally relevant, we shall be able to allow ourselves to universalize from Vere's case to other situations which may look very different in lots of ways, and so fend off triviality from the doctrine of universalizability.

It is worth preserving the doctrine of universalizability - and preserving it in a non-trivial form - because it models a strong intuition we have about the connection between action and reasons for action. If an action is performed for reasons, and if those reasons justify its performance, we do not suppose that the justifications could be unique or non-repeatable. However linked the justification may be to the details of the circumstance, the fact remains that the details (and I repeat that we are only interested in some of them) are not radically particular: they are general features which this particular circumstance (potentially among others) happens to instantiate. So the considerations telling in favour of action here could not have their force limited to this particular action. Someone who tried to particularize the impact of rationalization or justification in this way would in effect be withdrawing from the terms of the institution of justification itself. Winch is really making this withdrawal when he supposes that the morally relevant circumstances might be replicated while nevertheless supporting divergent practical judgments. The gap is then supposed to be filled - the slack is supposed to be taken up - by the agent's decision. But if the reasons for action (the morally relevant features) do not reach right up to that action, then they do not go any distance at all towards justifying it. The morally relevant considerations must either give an action its moral

character, or, if they do not, the action does not have a moral character. A bare decision could not 'take up the slack'; rather, it would take place in a simple moral void. And, as we have seen, it could do nothing itself to moralize that void.

At this point I register a difference from Wiggins' treatment of this matter. He writes:

In the presence of any case where neither finding is open to decisive criticism, and where the agent's actual finding is one we can make sense of (if only retrospectively) in the light of what the agent becomes, I am prepared to treat this finding as morally decisive of what was right for him. ([3], p.181)

But suppose that, had the agent made the other choice, he would have become something different but equally acceptable; why should we then say that, as things were, he made the right choice? The fact is that it didn't matter which choice he made, morally speaking. There was no moral right or wrong about it; his sheer decision could make no difference to that.

It will be said that the two senses of the phrase 'how things strike the agent' which I have separated out cannot be so easily divided from one another. Can it not be an aspect of (say) a man's sympathetic and generous character that he is disposed to admit a plea of extenuating circumstances to Budd's fatal act? But if we try to take this line here we can be faced with a dilemma. For is this disposition itself up for moral appraisal? If it is, then it must be content to take its place among the circumstances of the decision, being then judged morally relevant or not, and (if relevant) having its moral character assessed. But if the disposition is not itself up for appraisal, this can only be because the disposition is the

decision, or at least the impulse to make it, and so is consequential upon, rather than part of, the hurly-burly of moral debate in which the agent engages. (Of course the disposition is then up for moral appraisal in the sense in which the decision is up for it; but it is not a moral factor in the making of the decision: it is that decision-making process.) Winch seems to take this latter line: he can then only achieve divergence from the requirements of universalizability by supposing that the decision itself - and the disposition (in this sense) to make it - are somehow constitutive of its moral character. But we have rejected that claim.

We can now turn to the proposed solution of the puzzle. Placed in a similar circumstance to Vere's - that is, placed in a similar external circumstance - I might decide that I had to choose differently. I could make this decision without condemning Vere for his decision, only if I could conceive that something in his total circumstance (that is, his external circumstances combined with his character) made a morally relevant difference, sufficient to differentiate it from my total circumstance. Consequently I would have to find something in his or my character which made a moral difference. And it seems that I can achieve this conception, but only by employing the distinction for which I have argued, namely between the transcendental and the empirical components of the self. The self which acts is, as we have seen, the free (and hence transcendental) self. This allows us to detach it from the empirical self: I can now find some morally relevant difference between my empirical character and Vere's, without risking trivialization of the universalizability thesis by drawing the agents themselves into

the morally relevant circumstances. Were there no such difference between the transcendental and the empirical within the agent, trivialization would be inescapable.

It might be that in this particular case I should not be able to find any such difference-making feature of Vere's character; in that case I should have to be prepared to condemn his decision or change my own. But nothing, it seems to me, rules out that such a difference should reveal itself to inspection. Vere might decide, as I might not, for instance, that he could not live with himself if he allowed breach of the Mutiny Act to be seen to go without summary and confident punishment: this might be a morally relevant consideration (considered here as a datum of Vere's character, remember, and not as tantamount to an impulse to convict) if, for example, Vere's activities as captain would in his judgment be seriously impaired as a result of taking a more lenient course of action. Alternatively, Vere might notice a worrying tendency towards impetuosity in himself (this is actually hinted at by Melville), and decide to check it lest it precipitate undue haste in his deliberations.

Putting it generally, I might decide that my sort of character (and consequently the way I conducted my life) would be severely compromised by a summary conviction at this point; and/or I might allow that Vere's character would suffer damage if he did not convict immediately. These considerations would not be the only ones, of course, but they might play an important role in the deliberations. I am myself inclined to think that in a case as morally weighty as Budd's, it would probably not be in order to

place such emphasis on aspects of and effects on anyone's character; my own view is that a split judgment such as Winch favours is not really available in this particular case. But I am pursuing the principle of the thing, and in any case others might disagree. Certainly we should not fall into the temptation of supposing that for Vere to think about the effect on his character of adopting various courses of action must, of necessity, be selfish, given that the life of a man is in the balance; that need not be so (other lives have to be considered too).

No aspect of a man's character is in principle immune from being drawn into the morally relevant circumstances. That the agent is not thereby drawn into the circumstances - so generating triviality for the doctrine of universalizability - is guaranteed by its transcendental status, and by that alone. We noted earlier that there are two ways of taking the phrase 'how things strike the agent', one of which equates the phrase with 'what the agent is inclined to decide', and so locates this aspect of his subjectivity firmly on the side of the transcendental self, and the other of which cashes the phrase into specifications of aspects of the agent's character, so locating this aspect of his subjectivity on the side of the world. If we operate with a division between the empirical world and the transcendental self (with the empirical self, i.e. a man's character, being located on the side of the world), we can solve the puzzle we started with, allowing for decisions different from the ones we should ourselves make in (externally but not totally) similar circumstances.

FOOTNOTES

Chapter 1

1 Although the truth of this claim depends on the propositions' referring to rather than describing the cause and effect, i.e. that the items in question are referred to using, e.g., proper names: I explain this restriction below. I shall not always avoid the solecistic terminology; but it is fairly harmless shorthand which can always be replaced by a hygienic mode of expression. Cf. Maxwell, pp.149-50.

2 And so one should not take the undoubted truth that my perception of an object partly consists in the fact that '...if the thing had not been there, I should not even have seemed to perceive it.' (Strawson [4] p.51) as pointing to a relation of causal dependence, rather than 'logical'. Certainly the existence of a cup partially explains my seeing it; but I shall argue that the explanation here cannot be causal. Compare: the existence of a whole may be explained by the existence of the parts (along with a principle of combination); but the parts do not cause the whole.

3 'The Causal Theory Of Perception', reprinted in Warnock.

4 I discuss this relation, and the difficulties raised by it, below (1.5). The supervenience relation in question depends on a certain Cartesianism, or internalism, which I defend in 1.5.

5 A view associated with Davidson. See his 'Actions, Reasons and Causes' at p.14 of his [1].

6 For an account of the semantics of the 'NEC' predicate, see Wiggins [2], pp.108-11.

7 See further 3.2 below, and cf. Popper p.424-5.

8 This fallacy is committed by Maxwell p.158 (and making the statement of disposition more precise will not, as he supposes (p.160), rule out exceptions). But he concedes that where causal change over time is concerned, there is no logically necessary connection between cause and effect (since the process may be interrupted before the effect can materialize). Like Hume, Maxwell does not clearly distinguish between necessary connections between tokens and necessary connections between types. An argument for the existence of the latter is not an argument for the existence of the former. (In Hume's case: an argument against the existence of the former is not an argument against the existence of the latter.)

9 Cf. Wiggins [2], p.137.

10 Pears ([1], p.56ff) makes one of the mistakes I am castigating in his discussion of the parallel issue of whether objects of such feelings as amusement can also be (Humean) causes:

It is [no] good saying that hearing the remark or thinking about it cannot be regarded as causes of the person's amusement because his amusement cannot be separated from those

events. For it is only necessary to describe his amusement in a more abstract, general way that makes no reference to these events, and then it can be regarded as their contingent effect. (p.73)

The view that a state of amusement at something is such only under description surfaces here.

11 This position rules out anything like Kripke's doctrine of the necessity of origin, as David Wiggins has pointed out to me. But that doctrine is independently counterintuitive: can I not imagine a possible world in which I have different parents? Cf. Prior [1] ch. VII.

12 Cf. Goldman, p.265, on what he calls relation 'D': the relation of counterfactual dependence exemplified by someone's seeing an object in a mirror. Goldman draws an analytical connection between D and perception, and between another relation he calls 'R' (=an object's being in a suitable region of the subject's environment) and perception; he notes that neither of these relations is explicitly causal. But he wishes to incorporate causality in the modern conception of perception: this seems to me quite unwarranted: the modern conception is surely continuous with the older conception. What is true is that if we want to explain why lines of vision work in the way they do we shall have to descend to a level at which causation operates. Certainly a use Goldman claims for causality - to secure, for example, the result that frogs see - cannot be played by it: frogs were known to see long before their physiology was investigated, and on quite other (behavioural)

grounds.

13 An objection made by Evans at [2] p.200.

14 On this notion of elucidation see Wiggins [3] p.142 note 4, and p.188 note 4. Wiggins regards elucidation as an alternative to analysis, where the latter is unavailable. But the 'paradox of analysis' shows that, even where successful, analysis can in fact aim no higher than elucidation anyway. (See Blackburn, p.155, for a description of the paradox.)

15 I should indicate - if it is not already clear - what my attitude to Nozick's treatment of knowledge and scepticism is: in brief, Nozick is right to say that we do not 'track' the truth/falsity of the sceptical hypotheses, but wrong in his claim (surely absurd) that knowledge is not closed under known logical consequence. (Wright, [2] p.445, oddly suggests that Nozick's principle 'must be sometimes valid...if logical inference is ever to be a source of new knowledge.' But what we require for the truth of the protasis is that knowledge should not be closed under unknown logical consequence - as indeed it is not.) What we now say about our epistemic relation to such propositions as

(M) There is a piece of paper in front of me depends on whether we take an internalist or an externalist approach. The internalist will say that I can know (M), and know that I know it etc, more or less whenever it sensibly seems to me to be the case; but that (M) does not entail

(P) There is a material world
since (M) relates to a provisional world, and (P) to the (ultimate) world. So Nozick's sceptic can't contrapose on my ignorance of the truth of (P). The externalist will say that (M) and (P) concern the same world, so that there is an entailment from (M) to (P), and consequently that the sceptic can contrapose, given the (undeniable) fact that we might be mistaken in our confidence in the truth of (P). But the sceptic cannot upset our right to come out with propositions like (M) and (P), and that is all that really matters (the striking and significant fact here being that the sceptic might be wrong, not that he might be right). Again, I do not detect a metaphysical or epistemological superiority of one of these approaches over the other in their responses to Nozick's sceptic.

Epistemological scepticisms - I do not know that I am not a brain in a vat, dreaming, that there is a material world, that the world has lasted for more than an hour etc. - are not refutable: we do indeed fail to know all of these things. But they do not and cannot challenge our right to work on the assumption that these and such things are true (to work with a concept of materiality, the past etc.). Epistemological scepticisms cannot be upgraded to (what really would be frightening) metaphysical scepticisms which challenged our right to operate with such concepts at all.

16 I make no claim concerning the extent to which the view presented in the text is Wittgensteinian. In one important detail it is not: I speak of an internal, constitutive connection between appearances and states of the world which those appearances

represent. This connection is not, contrary to what Wittgenstein is prone to assert, a matter of grammar or convention. The criteriologist's position is to be found (along with much irrelevant and unwanted embellishment) in Baker. A saner account is to be found in Wright [1], and Albritton. But it will be clear from the text that I dissent from Albritton's postscript, in which he suggests that, after all, no sort of behaviour - under ordinary, non-question begging descriptions - enjoys necessary connections with inner states. The point is precisely that for much - perhaps all - behaviour non-question-begging descriptions are not available. I discuss criteria in relation to the other minds problem further in chapter 5. Here the point is that no faithful description of an appearance of rain can possibly omit the fact that the appearance in question is one of rain.

17 The best version of the disjunctivist position is presented by McDowell in [3], although he partly relies on earlier work by Hinton.

18 McDowell, [3] p.474, suggests that the criteriologist's way with criteria 'undermines the very idea of an appearance having as its content that things are thus and so in the world 'beyond' appearances (as we would have to put it)'. But it is exactly because we don't have to put it so that the undermining does not take place (and hence that a metaphysical scepticism - see above note 15 - does not arise for the criteriologist). The internalist's external or real world is not beyond appearances (that is to bring in an

irrelevant piece of externalist rhetoric); rather it is the content of the logical combination of (corroborable) appearances. It is easy to become so thirled to a particular picture that one invests it with a metaphysical power which it simply does not have; hence the attempt to embarrass phenomenologists of the 'logical' sort - see below in the text - with the question: but what about the beyond? The notion of 'beyond appearance' can figure quite happily in the realist's model (appearance then playing as mere appearance); but it has no place in the phenomenologist's.

Slightly later, McDowell argues that 'we find the match in content [between the appearance that it is raining and the knowledge that it is raining] intelligible only because we do not conceive of the objects of experience as in general falling short of the meteorological facts.' But the criterialist's position is that we find the match intelligible because the relevant criteria do not in general get defeated. What is the difference (detruding McDowell's 'only') but one of notation? (Of course the crucial difference remains that the disjunctivist does not describe the phenomenology of experience correctly.)

19 Related to, but not identical with, the doctrine often called 'linguistic phenomenalism' (see, for example, Mackie [1]). But only related to, because language does not, at least in the first instance, have anything to do with it. Mackie's elaboration of the doctrine - with appearances featuring as the internal accusatives of perceptions (pp.117-9) - is rather close to my account. But Mackie thinks that this phenomenalism gets confused between the internal

and external accusatives of perceptions. Once the argument from illusion has done its work in reifying all the internal accusatives of my perceptions, 'I am enclosed in a circle of mind-dependent objects: how can I ever break out?' (p.120). Of course I neither can nor need to. The mistake was made earlier in the tacit assumption that if the reified object of my hallucination is non-material, the same goes for all the (internal) objects of my perceptions (so that I am cut off from the material world). But, clearly, the objects of my perceptions will not be non-material if those perceptions are veridical. For in that case the internal objects will be external objects. External or material objects just are, on this picture, the objects of all actual and possible perceptions of them. The subject is confronted at least with intentional objects, which he normally takes to be real objects. If his assumption is correct, his intentional objects will be 'external' or real objects - objects capable of figuring as the content of indefinitely many experiences. Our phenomenalism does not regard mere appearances but rather objects of appearances (or: appearances with objective content) as the basic fabric of the world: see further on this 1.4.

20 Here and elsewhere all references to Kant, unless otherwise specified, will be to the Critique of Pure Reason in the pagination of the 1781 and 1787 editions. Otherwise, reference will be made to the pagination of the edition published by the Königliche Preussische Akademie der Wissenschaft, abbreviated 'Ak'.

21 Cf. Ayer [1] pp.87-9.

22 Cf. Quinton [1] passim, especially p.68; Wilkes, p.96.

23 Mackie ([2], ch.2), following Locke, makes this error.

Indeed, as soon as one raises the question of resemblance between veridical appearance and reality - as opposed to identity - one has stepped into the incoherence castigated in the text. See McDowell's criticisms of Mackie in [4], especially pp.114-5 with notes 14 and 15.

24 Especially the considerations telling against the possibility of type identities between the mental and the physical adduced by Davidson, most notably in his 'Mental Events' (in his [1]). In fact the diverging commitments of mental and physical characterizations, which Davidson rightly stresses, count as much against token-identity as against type-identity theories: for the ascription of a particular mental event or state M might, in context and under constraints of rationality, compel interpreters to ascribe another mental event or state, say N. But if M supervenes on the physical event or state P, and N on physical event or state Q, it follows that the neurophysiologist who ascribes P to the subject will be forced additionally to ascribe Q to him. But then, absurdly, it ought to be possible to do neurophysiology a priori: the physical is being constrained by considerations proper only to the mental. ('Q' functions here as a descriptive name, but that does not affect the metaphysics of the point: that the psychological a priori is being allowed illicitly to constrain what ought to be no better than the

physical a posteriori.) Davidson ought to have been alerted to the unacceptability of the token-identity theory by his own attacks (see various papers in his [2]) on the scheme-given dualism. For his claim that events are mental or physical 'only under a description' lands him precisely in an ontology of bare particulars awaiting schematization.

The other main group of considerations relate to qualia, and the apparently evident conceivability that qualitative mental states might be realised anyhow (see especially Kripke [1]). Functionalists such as Shoemaker (see here Shoemaker and Swinburne, pp.106-7) incline to materialism for such reasons as that 'in the absence of compelling evidence that mental states are realised non-physically we ought to believe that they are realised physically.' But only dogma can carry us further to the materialist's assertion that they must be so realised. Not that they can be realised in an immaterial substance: it can only queer the dualist's pitch to suggest that this is the alternative, rather than that they are not realised at all. Talk of immaterial substances can be harmless if it is construed as a facon de parler for the persisting self; but dualists should not present such talk as if it had any explanatory value.

25 One aspect of the curious nature of supervenience is brought out by Blackburn (p.182ff): it has a commitment to what he calls a ban on mixed possible worlds. That is, if A is a supervenient property (and hence essentially supervenient), and if B is its supervenience-base in the designated world, then we can imagine worlds in which everything B is also A (as in the designated world);

we can further imagine worlds in which nothing B is also A (A supervening on, say, C in those worlds, or simply not existing there). But we cannot imagine mixed worlds, i.e. worlds in which some B-things are also A, and some are not. Why not? Alternatively, if just one B-thing is also A, then everything B in that world has to be A too. And if one B-thing is not-A, then everything B in that world has to be not-A too. But why? (Blackburn's own claim - that a projectivist metaphysic is better able to cope with this puzzle than a realist one - is the very reverse of the truth. At least the realist can present us with the brute puzzling fact about supervenience and leave us to make of it what we can. The anti-realist, in suggesting that the ban on mixed worlds was in some way our decision, makes the puzzle even deeper. Why on earth should we have made that decision? Mystery on God's side is all very well; mystery of our own making is doubly mysterious.)

26 In the terminology which I shall explain and employ in chapter 3, the internalist's claim must be formulated in A-series terms rather than B-series terms.

Chapter 2

- 1 See McGinn [2] p.5, Evans [1] p.95, McDowell [4] p.111-2.
- 2 Cf. Evans [1] p.109, note 43.
- 3 Cf. Strawson [1] ch.2.
- 4 This is clear also from the economic analogy which Evans deploys. The value of the pound is reducible (not to any one exchange rate but) to the totality of exchange rates: it is not something over and above that totality. Similarly, for Hero there is supposed to be an overall reducibility of statements locating sounds at positions to statements relating sounds to one another in an essentially phenomenalist way (but without the objectivist commitments of the phenomenism of the last chapter: Evans' phenomenism can be captured, as mine could not, purely non-conceptually).
- 5 The terminology is due to McDowell; the associated thesis was discussed by him in lectures given at Oxford during 1985.
- 6 [2] p.142.
- 7 Of course the categoricity here does not preclude

dispositionality which is not mere dispositionality (i.e. dispositionality without categoricity). We can, for example, acknowledge that secondary qualities are dispositional in the sense that they present different appearances to perceivers in different circumstances. But there is no connection (contra Dummett [2], p.31) between this dispositionality and the physicist's account of the phenomena which secondary qualities supervene on, or are correlated with. To say that it is part of what it is to be yellow that a yellow object look differently according to circumstances (including perceivers' sensory equipment) is in no way to set the stage for 'an identification of colour properties with physical properties not in any way characterized in terms of colour vision.' Such physical properties are not essentially observational, unlike colour properties. There is no reason at all to suppose, in advance, that the observational should consist in what is not observational, and perhaps not even observable. I shall argue in 2.3 below that the categorical status of secondary qualities does not conflict with another sense in which they are dispositional, that captured in the formulation: x is red = x is such as to look red.

8 [2] ch.1 especially pp.15-6.

9 The two aspects of this empiricist analysis are also to be found in McGinn (see [2] pp.115, 126-7). McGinn, like Evans, wants the primary to objectivize experience.

10 Evans no doubt derives his assimilation of representational

and non-representational experience from Wittgenstein, whose discussion of privacy is vitiated by the failure to distinguish between them. Observe, for example, the silent substitution of 'red' for 'pain' at PI 1.271-2. It should be noted that the term 'non-representational' is somewhat unsatisfactory as a characterization of the sort of experience in question: non-representational experience is not entirely non-representational; rather it is not entirely representational. When I run my finger along a table-top, I feel both myself, and the table-top: I experience both world and self. At least that is one way of putting it. Whereas when I see the table-top, I only experience the world. But, with this caution, I shall continue to use the traditional terminology.

11 Cf. McGinn [2] p.14, and Campbell.

12 See, e.g., [2] ch.8.

13 See Blackburn's Spreading The Word passim, and elsewhere in his writings. Cf. also Wollheim for a discussion of projectivism in relation to aesthetic value. Wollheim suggests that there is a distinction between works of art and works of nature of the following sort: there are standards of correctness for projections of value onto works of art, but not works of nature (p.12). But it is incoherent to talk of (in)correct projection. For a work of art (or anything) to be such that certain projections of value onto it are (in)correct, is just for those values to be there (or not) in

the work: so projection is not in question; what it is for a value to be there in the work is that the beholder is under a constraint to respond in a certain way (see here McDowell [4], p.119). Wollheim ought to express his point by saying that while a projectivist treatment is appropriate to value in works of nature, what is required for value in works of art is objectivism. Thus emended, the position still seems to me to be inherently unstable towards (unrestricted) objectivism: for how could a painting of a landscape (say) have the real, unprojected quality of serenity, if the original scene did not in some way possess it? (None of this goes to show that it is not possible to make projections of value onto things or artefacts, only that not all values can be construed as the projectivist wishes.)

14 Although in one place (Principles I.III) Berkeley indicates the kind of view I favour: he offers the beginnings of a subjunctive conditional analysis of what it is for a table to exist in his study.

15 Cf. Dummett [2], p.12, McGinn [2] passim. For the idea that such primary qualities as the squareness of a table are experiential, see Kant, Prolegomena, First Part, Remark II (Ak. 288-90).

Chapter 3

- 1 For an account, see Dummett [1] pp.351-7, and Mellor ch.6.
- 2 At pp.107-8.
- 3 A thesis argued for with ingenuity by Mellor in his Real Time.
- 4 In his unpublished 'Demonstratives'.
- 5 I understand Frege's notion of sense non-descriptively, according to the interpretation to be found in ch.1 of Evans [2], or in McDowell's [5].
- 6 See Frege's 'Der Gedanke' and Evans [2] ch.6 appendix.
- 7 See the paper of that name, reprinted in his [3].
- 8 Of course, if the action to be explained is characterized in A-series terms ('Why do I get up now?'), then only the citation of an A-fact will satisfy ('The meeting is at 3 p.m., and 3 p.m. is now'); but the engagement of the A-series with rationalization is extraneous or, as I shall put it, transcendental. Nothing of importance to the rationalization is omitted if action and explanation are specified entirely in B-terms.

9 Cf. Sellars [1], especially sections I-IX.

10 If we insist that the formulation 'X is past from the point of view of Y' is an A-formulation because it embodies the A-notion of the moving present, we can answer McTaggart's famous argument for the unreality of A-time (and hence of time itself). McTaggart argued that the supposition of the reality of A-time leads to the contradiction of each event having three mutually incompatible properties - of being past, present and future. The answer to this argument is to point out that the possession of these properties is relative to different times. If an event is present at a time t it is future at times earlier than t and past at times later than t . This is not an exclusively B-formulation (contra Horwich ch.2), because the notions of past, present and future incorporate the A-idea of the movement of the present moment. But it does presuppose an integration of the A-series with a B-series. McTaggart's regress only arises when the relativization to times of A-characterizations itself has to be made in purely A-terms. If we relativize to dates, the regress disappears.

11 By Grünbaum in loc. cit., note 2.

12 [1] pp.321,324.

13 See [2] e.g. pp.14-5, pp.145-6; [6] passim.

14 This does not introduce transcendental objects in Kant's

sense, because an object comes packaged with no more than two possible sets of causal dispositions. So there is no notion at work here of an object prior to all empirical determinations. Objects (on this picture) are empirically determined up to (but not including) the establishment of the time-direction.

15 See the description of the case and discussion at Mackie [3] p.183ff.

16 See [3] p.164ff.

17 A point ignored by Lewis in his treatment of the question. Lewis asserts (p.474) that joint effects overdetermine their common cause, but this is incorrect. An effect E_1 determines its cause C only if all its joint effects $E_2 \dots E_n$ co-occur (holding other processes and causal laws fixed in so far as possible). C would not have occurred, in those circumstances, if only E_1 and not also $E_2 \dots E_n$ were going subsequently to occur.

18 Horwich (p.49) thinks that the question whether or not time has a direction is this question: is there a physically significant difference between past and future orientations? But what would an answer to this question tell us about time? Any difference between the orientations (for example if the laws covering one direction are simpler) is necessarily neutral as to which direction is past and which is future. This is not a matter of arbitrary labelling; the difference between past and future is not a matter of terminology,

but a deep ontological difference.

19 Lewis (p.460) expresses a similar point to this, but with an epistemic slant: 'it is less certain that our world is indeterministic than that there is an asymmetry between an open future and a fixed past.' But the point is not how certain we are. We might be as certain as you please about quantum indeterminism; but we could never jump the truth of that indeterminism up from its contingent status. And nothing necessary - such as the anisotropy of time - can be constituted by anything contingent.

20 Cf. Lucas, who discusses the question whether this view of time is challenged by the Special Theory of Relativity. For my purposes it doesn't matter whether there is such a thing as absolute simultaneity or not. The characterization of the difference between past and future can be supplemented by relativization to frame of reference without affecting its validity, so indexed. The reference-frame can, if you like, be indexed to a particular agent; it will remain true that, for that agent, there is an ontological difference between his future and his past. The argument in the text that agents cannot enjoy different pasts from each other should receive the tacit gloss: cannot, relative to some to some (any) frame of reference. Whether or not there is a physically preferred frame of reference can be left open.

21 Russell [2] pp.193-4, Ayer [1] pp.164-6.

22 Cf. Denyer section 43, Gale ch.6.

23 As Mackie puts it in his discussion of Newcomb's Paradox (which I assume familiarity with): 'if we...let the player know that he is under hypnotic control, but still assume that he cannot break free from it, and knows this, he may reasonably hope that he has been directed to take the closed box only, and be pleased if he finds out that this is in fact so. But this does not make it reasonable for him to choose this strategy: on the present hypothesis, no such choice is open to him.' ([4] p.151). What goes for hypnotic control goes also for a psychologist's reliable prediction on the basis of character (p.152).

On the other hand, if the player takes himself to be free then he surely cannot suppose that the experimenter has predicted his action, or even predicted his likely action (the supposition of genuine freedom rules even this out), so that he cannot suppose that what he in fact does (which box he opens) makes any difference to the presence or otherwise of money in the sealed box. Freedom to choose between two courses of action requires nothing less than a probability assignment to each possible action, up to the moment of decision, of exactly one half.

The situation which Mackie and I are agreed in disqualifying is that in which he takes himself to be free (or: he is free) and in which in addition he takes the experimenter's predictions to be (or: they are) reliable indicators of how he will act. If he holds the first (or: if it obtains), he ought not to hold the second (or: the second cannot obtain) and vice versa.

Dummett, [3] p.165, complains that the issue of whether or not it is rational to take both boxes is not connected to the free will/determinism issue, and argues that whatever the agent's views concerning his freedom or lack of it, at the end of the day he still has to choose, and it is open to him to choose more or less rationally. But these contentions are exactly tantamount to a statement to the effect that the agent is free. It is nonsense to say: never mind about freedom, the agent still has to choose. If he has to choose, then he can choose, and the issue of freedom has been settled. His having to choose entails that he is free. And if he is free, then the psychologist cannot have predicted his choice. For his freedom requires the equal likelihood in advance of each choice (cf. Prior [1] ch. IV).

Of course we can and do predict each other's actions - many of them as good candidates for the status of free actions as we have - and it seems quite conceivable (although it would be remarkable) that a psychologist should achieve the degree of predictive success required by the terms of Newcomb's Paradox. What this shows, however, is not that it is therefore unproblematic that one can both be free and judge it more rational to choose one box only, but that we have two mutually exclusive ways of thinking about human action, one which embraces freedom, and one which embraces something which excludes freedom, namely the explicability and predictability of action. We cannot simultaneously and consistently think of action in both these ways. The paradox arises from the attempt to do so. (I explore these inconsistent commitments below, 3.5 and 3.7.) But the agent, in thinking of his own actions, cannot help thinking that he

is free (3.6); so Mackie gives the right answer on the question of what the chooser in the Paradox ought rationally to do.

24 Cf. Wiggins [4], pt.II

25 Contrast this scepticism with the epistemological scepticisms which, as we saw in chapter 1, can always arise for subjects of experience. Epistemological scepticisms are - in contradistinction to metaphysical scepticisms - not very serious. But it is a metaphysical scepticism (inability to establish the time-direction, and hence lack of entitlement to the concept of a temporal world) which hangs over the compatibilist. It is not just that the spectator-agent cannot know which direction is the future and which the past; more than that, he cannot form a conception of what that difference would consist in.

26 Cf. Ayer [1], p.152.

27 Cf. Whitrow, pp.74-7.

28 Strawson [2], p.137ff.

29 Mackie [3], p.131. Cf. also Flew in Flew and Vesey, p.135.

30 Mackie [3], ch.2.

31 In his Freedom and Belief.

32 Cf. Descartes Fourth Meditation:

For in order that I should be free it is not necessary that I should be indifferent as to the choice of one or the other of two contraries; but contrariwise the more I lean to the one - whether I recognize clearly that the reasons of the good and true are to be found in it, or whether God so disposes my inward thought - the more freely do I choose and embrace it...This indifference which I feel, when I am not swayed to one side rather than to the other by lack of reasons, is the lowest grade of liberty, and rather evinces a lack or negation in knowledge than a perfection of will. (HR p.175)

33 But Kant's argument here is defective, relying as it does on the spurious need to support conditioned causal chains with something unconditioned. Cf. Strawson [2], p.208.

34 As Strawson famously argued in his [3].

35 This consideration has been regarded by many - in my view rightly - as decisive against the possibility of travel into the past. (The possibility of 'travel into the future' which is allowed for by the Special Theory of Relativity - i.e. the possibility of travelling quickly and thereby ageing slowly - is not affected by the considerations mentioned in the text.) Horwich suggests that the following scenario is coherent:

I go back three years in time, return to my house, and discover a person whom I take to be an earlier version of myself sitting in the living room....I am absolutely confident that within the last ten years or so of my life I have never been punched on the nose. However, despite all this I feel, and presumably am, free to punch this person on the nose. And whether or not I do so, no contradiction will be engendered. If I do not, then all is well with my faculties. If I do punch him, this means that things were

not as I remembered them. (p.117)

Even waiving the difficulty involved in the supposition that I encounter an earlier version of myself (a difficulty itself fatal to the hypothesis of backwards time-travel), the passage makes the intolerable assumption that until, in the envisaged scenario, I act (i.e. until I punch or not), it is not yet settled whether my apparent memories that I have never been punched on the nose during the last ten years are veridical or not. That gets settled, it seems, only when I decide whether or not to throw the punch at my earlier self (at least it gets settled if I decide to do so).

But this account is committed to an anti-realism about the past which is implausible, and which I go on to reject in the text. Before I decide whether or not to throw the punch, it is fixed and determinate what actually happened at the earlier time, and whether my apparent memories are genuine memories or not. (Of course I may not know which of these is the case). And then it is a simple consequence of realism about the past that I cannot, at the moment of decision, regard myself as free to act either way. If my apparent memories are veridical, I cannot punch; if they are not, then either I must punch or I cannot, depending on what the content or a veridical memory (in principle available) would be. Whichever of these three possibilities is realised (and it is already, before I act, determinate which one is realised), I am not now in a position to come to a free decision as to how to act.

It might be thought that we could salvage travel into the past if we provided that some areas of action should be closed off from the traveller's access. But by the ontological parity of all past

events (see below in the text) the traveller would be excluded from initiating free action in any area. And that is just to say that he would be excluded from action simpliciter. So travel into the past is not possible for an agent. There is, as I hinted above, a more general argument against the possibility of time-travel for anything, to the effect that such a possibility would require, absurdly, the identity of two different times. (Or, to put it differently, it would require the simultaneous movement of the traveller into the future and either into the past or into a distinct future.)

36 For a description of recursion in memory, see Evans [2], 5.2. I avail myself of this structure since it is familiar; but actually I do not think anything of epistemological importance turns on whether we regard memory as having a recursive structure, or as giving the subject direct access to the past (he literally sees those past events). For this treatment of memory, see Russell [1], pp.26,66. The objection usually deployed against Russell - that he accords the past a ghostly unreachable existence in the present - is spurious. It simply fails to take seriously the special nature of perception which memory is being cast as: namely that it is perception, precisely, of the past. If we followed Russell's construal, the argument against anti-realism would proceed via just one fact (the representational nature of memory experience), not two, as my version in the text has it. On Russell's model it is also easy to see how it is possible to reject the inclusion of causality in one's account of memory (contra Martin and Deutscher), in a move

parallel to that made in chapter 1 for the case of perception. But the recursive structure does not necessitate a causal account either: the perception which is retained and recycled when the original scene is remembered is not causally related to its environment.

37 Dummett's purported cases of backwards causation in arts. cit. (note 12) depend on suspending the epistemic parity between the known past and the unknown past. They depend on taking the unknown past as not yet fixed. On my argument in the text, that is tantamount to taking the unknown present as not yet fixed; and this can only be done by falsifying the content of representational experience. There is no extra difficulty about the past.

38 As Dummett has made clear. See e.g. his [3] at pp.147-50.

39 So G.Strawson, pp.133ff.

40 If we confine ourselves to the actual world it is at least conceivable that no functional difference would be discernible between, say, someone who saw the world as we do and someone who saw it under red-green inversion (or between two subjects each having monochromatic vision but in different colours. Cf. McGinn [1], pp.34-6). Of course we can devise counterfactual situations to enable us to draw a behavioural distinction. One such is the following. Consider X, who has normal vision, and Y, whose vision is red-green inverted. This difference partly consists in the fact that

were Y to enjoy a colour-experience qualitatively identical to X's when X looks at a British post-box, he (Y) would group that experience with those he obtains when he looks at limes, grass etc. (whereas X groups that experience with those he obtains when he looks at blood, ripe tomatoes etc. Cf. Pitcher p.205). But there is an objection to the functionalist's availing himself of this behavioural 'manifestation' of inversion. For the functionalist cannot distinguish between the envisaged counterfactual scenario and its symmetrical opposite, i.e. where X counterfactually enjoys a colour-experience qualitatively identical to Y's when Y looks at a post-box, in which case he would group his experience with those obtained when he looks at limes, grass etc. (whereas Y groups his experience with those obtained when he looks at blood, ripe tomatoes etc). Both counterfactual scenarios - X seeing an object as Y sees it, all other experiences being held constant, and Y seeing that object as X sees it, all other experience being held constant - have the same behavioural (including linguistic) manifestations.

This objection covers the functionalist's attempts to deal not only with the case of intersubjective inversion, as above, but also with the case of intrasubjective inversion (as when I wake up one morning to find my colour-experience red-green inverted). In the intrasubjective case a functionalist cannot distinguish the scenario in which I wake up having suffered a red-green inversion in one direction from the case in which I wake up having suffered it in the other direction: in both cases, the only manifestation of the change which the functionalist can pick up are such utterances as 'limes and post-boxes have exchanged their colours'; and obviously any such

utterance will be available in both sorts of inversion, and so will not tell us which one has occurred.

(Note that the argument depends on there being no other way (relating, say, to differing powers of discrimination) of distinguishing between red vision and green vision than the purely phenomenological: it is then argued that the phenomenological need have no functional manifestation. The position just requires the conceivability of inversion. It may in fact be the case that there is no 1-1 mapping of the colour spectrum onto itself, other than identity, such that topological relations are preserved - in the sense of not affecting the discriminatory abilities of perceivers (see here Harrison, part 4); the point is that we can imagine that there could be cases of inversion which do preserve discriminatory abilities of perceivers. On these issues see especially Shoemaker, who is however too sanguine about the functionalist's chances of analyzing the possibility of intrasubjective inversion.)

Chapter 4

1 On this locution see below, 5.2. Cf. Strawson [1] ch.1 on the fundamental status of material objects.

2 Cf. Quinton [2] pp.12-5, Strawson [1] p.20.

3 Cf. Strawson [1] ch.4.

4 Cf. Quinton [2] pp.17-8, Strawson [1] pp.36-8.

5 See the lucid account in Davies, ch.5, with further references. Especially significant are McDowell [1] and [2], and Evans [2] *passim*.

6 See especially [2] ch.10.

7 Russell's own restriction of the objects of acquaintance to sense-data can be accepted (contra McDowell, [5]) as long as we construe sense-data in the spirit of the phenomenalist alternative of 1.3.

8 As Ayer puts it ([2], p.161): 'an underlying motive for getting rid of singular terms is the desire to dispense with the category of substance [=particularity].' But he then adds, incorrectly: 'To show that all the work of singular terms could be done by predicates would be a way of vindicating Berkeley's thesis

that things are bundles of qualities.' This is incorrect, because as soon as we collect qualities into bundles (in the sense in which a distinction is drawable between two numerically different bundles of the same qualities), we have substances.

9 See Walker, ch.IV for a discussion and elucidation of the phenomenon of incongruous counterparts.

10 Julius Caesar I.2.52-3.

Chapter 5

1 [4], p.36.

2 A point first brought to prominence by Castaneda. Cf. also Perry.

3 At p.328ff.

4 Williams [2], ch.3.

5 To Regius, December 1641 (K, p.121ff), January 1642 (K, p.126ff).

6 [3] p.195.

7 Cf. McDowell [4]. 'Constitutively such as to provoke laughter' would be too strong.

8 Although there is room for a certain relativism here. But once the relativizing has been carried out, it is still true that within a particular group (collected on the basis of propensity to find certain kinds of thing funny) jokes may fail to be appreciated, and laughter may be directed at what is not (by the group standards) funny.

9 Cf. Wiggins [2] p.159 footnote 13.

10 So we can say more about the identity of the soul over time than - what is only true if it is read in such a way as to make it utterly trivial and uninformative - that it involves the persistence of the same immaterial soul-substance. We can say something of what that persistence consists in. Swinburne's dualist - with whom I have a lot of sympathy - mislocates the role that memory plays here (see p.9 of Shoemaker and Swinburne). It does not supply us with a serviceable criterion for personal identity, since in hard cases, where our normal bodily criteria (they are criteria of a defeasible sort) are waived, we should have no reliable way of distinguishing apparent from veridical memory (cf. Penelhum, ch.5-6). But the fact that we cannot get a reliable test out of memory is irrelevant to its unimpeachable constitutive status (which it shares only with intention) for personal identity over time.

11 Failure to make the crucial distinction between memory and intention on the one hand, and all other psychological attitudes on the other, is part of what vitiates Parfit's recent attempt to discredit the notion of 'absolute' identity. Even if Parfit's hypothesis of a kind of 'memory' ('quasi-memory') that did not presuppose personal identity were obviously coherent - which I am very far from allowing - nonetheless our practice of operating with a concept of absolute identity would just so far remain unimpugned; since we never in fact encounter quasi-memories. All the memories we have presuppose absolute personal identity. (Parfit is followed in his erroneous conflation of identity-presupposing psychological

attitudes with non-identity-presupposing ones by some of his commentators: see, e.g., Sprigge at p.49.) Parfit's clumsiness with the psychological - especially his failure to make the above crucial distinction - lies at the root of the intolerable conclusions he reaches in part III of his book. That the identity of the transcendental self is indifferent to changes in such psychological attitudes as (world-directed) beliefs and desires is, I believe, the right conclusion to draw from Williams' famous thought-experiments in his 'The Self and the Future' ([1] pp. 46-63).

It is true that apparently coherent thought-experiments can be devised in which absolute identity breaks down, and a desire for survival without identity looks more rational than a desire for strict identity (for a good case, see Shoemaker in Swinburne and Shoemaker at p.119ff). But that does not go to show that identity is not what matters in the ordinary case. One might say: if strict identity is available, it matters supremely; in certain fantasy cases (involving, e.g., the duplication of my body) where it is not available, some kinds of survival would matter instead. But here we must insist that survival of the sort which preserves memory-like and intention-like states (q-memories and q-intentions) - i.e. states which would be memories and intentions if there were no duplicates of my body - is the only sort of survival which matters in such conditions. Similarity of character, appearance etc. does not matter at all: it has nothing to do with either identity or its tolerable surrogates. But a major query still hangs over the coherence of such thought-experiments.

12 Cf. Swinburne in Shoemaker and Swinburne. p. 58-9, who notes that the main difficulty facing a dualist theory is to explain why bodily continuity is regarded by us as having anything at all evidentially to do with personal identity, and suggests that the reason is to be found not indeed in any necessary connection, but in our observation of this regularity: same person iff same body. But there ought to be an explanation of this regularity; and we do not know (within dualism) what it is.

13 A 363 footnote a. Cf. Strawson [2] p.168.

14 The nub of this objection was well put by Wittgenstein: If someone says 'I have a body', he can be asked 'Who is speaking here with this mouth?' (OC, 244).

15 Contra, e.g., Pears [2], p.170, following Hume.

16 See Wiggins [1] on the undesirability of rejecting de re necessity, on the availability of which this claim depends. And cf. 1.1. But the necessity here runs in the opposite direction: both representational and non-representational experiences are essentially of what they are of; objects of non-representational experiences are additionally essentially owned by the experiences which are of them.

17 Cf. Locke 2.27.1 and Kripke [1], for the view that an object essentially possesses its origin; a view which, however, I rejected

above (1.1).

18 [2] p.80.

19 Wright himself offers the private linguist another riposte to the objection: the rule-governed nature of the private linguist's practice may be simply ungrounded; and none the worse for that, since many of our beliefs (the so-called 'framework' ones, discussed by Wittgenstein in OC) are equally ungrounded, and so incapable of being justified. But lack of justification in these case does not mean that the practices or beliefs are unjustified. Cf. Wittgenstein PI 1.289: to use a rule without justification does not mean to use it wrongfully. (For the correct translation of this passage, see the excellent remarks of Kripke [2], p.74 note 63.)

It might be worth appending a discussion of Wright's own preferred construal of the Private Language Argument in his [3], since it is a recent and sophisticated attempt to run the seems right/is right interpretation of the argument. Wright makes two claims:

- (1) There is a connection between the seems right/is right distinction and factuality, i.e. the distinction has to be drawable for a range of purported statements if they are to count as factual.
- (2) The private linguist cannot make the distinction for 'statements' about private objects.

The argument for (1) proceeds as follows:

- (A) The Cartesian claims that
 - (a) X believes what 'P' expresses and
 - (b) What 'P' expresses is trueare always co-assertible (because they mutually entail one

another) where the domain of discourse is X's sensations.

(B) But the Cartesian must be using the operators in (a) and (b) in their usual sense.

(C) In their usual sense, where 'P' is a substituent about which X has no special authority, the mutual entailment fails (fallibility enters).

(D) Hence the Cartesian must supply a difference in content between (a) and (b) where there is special authority, other than the difference accruing to the operators in virtue of the way they come apart in contexts where X has no special authority.

This argument is fallacious. The Cartesian can reply that (a) and (b) do not differ in informational content in the special case where X's sensations constitute the domain, and 'P' is a sentence like 'I am in pain'. The operators in (a) and (b) have their usual senses, and hence do differ in sense, because in other contexts and over other domains (a) and (b) can differ in informational content (can fail of co-assertibility). But there is no reason why that difference in sense should manifest itself additionally in the special case; there is no warrant for extending the fallibility requirement from the non-sensational to the sensational domain. In fact this seems to me to be right whether or not there can be such a thing as a private language. At least in the case of X's own pain, (a) and (b) cannot come apart. So Wright's view also fails to meet the plausible constraint on the cogency of any argument against private language, that going public should make the difference: public or private, (a) and (b) are always co-assertible, where the domain of discourse is X's own sensation.

Claim (2) is that the private linguist cannot make the seems right/is right distinction; if I am correct we should accept this claim - at least for certain central cases of 'inner' experience, and for judgments at an instant - but not regard it as damaging,

because it applies to the public as well as the private linguist.

In fact the constraint on cogency is doubly breached, because an analogous difficulty to the one canvassed above faces the entire linguistic community. That community cannot make the relevant (a)/(b) distinction where what substitutes for 'P' is a 'framework' belief (suitably adjusted to render it sceptic-proof, e.g. the belief that the world seems to have existed for more than five minutes), the sort of belief adoption of which is partly constitutive of rationality in our world. Wright (p.236) argues that the community can envisage (a) and (b) coming apart by imagining an outsider who sees that the community errs in one of its framework beliefs. But how could there be someone who was simultaneously an observer of and an outsider to such beliefs? Anyone who got as far as understanding our beliefs in such fundamental matters would have to share them. There is no hint of verificationism here, no supposition that there could not be forms of rationality in principle inaccessible to us. But an outsider who observes and understands us is not such an inaccessible centre of rationality. The argument is just that to understand us you must share our basic beliefs; and that seems highly plausible. (If you didn't share them, how on earth could you get to know what they were?) So the community has beliefs for which it cannot make the (a)/(b) distinction just as much as the private linguist has.

So far, we have just been considering the problem of making a synchronic distinction between (a) and (b). But the Wittgensteinian polemic is mainly directed towards the diachronic case, i.e. involving substitutions such as 'this is sensation S again'. Now

arguably it is correct to insist that the subject must be able to distinguish between its seeming to him that sensation S is recurring, and its actually being the case that it is recurring. He must be prepared to recognize the possibility that he can make mistakes in such judgments, and we need a warrant for supposing that he can do so. But what stands in the way of his recognition that he can make diachronic mistakes about the phenomenology of his experiences? (He is, in Wright's terms, a fallibilist about his experiences.) Wright claims (p.238) that 'phenomenologically there simply is no difference' between the sensation's undergoing change, and one's changing one's mind about which sensation it is. But this seems to be just false. There is certainly a phenomenological difference between on the one hand the sensation's noticeably remaining the same and one's changing one's mind about which sensation it is, and on the other hand its undergoing a noticed change (cf. the remarks on Hero's phenomenology of receptivity in 2.2). The subject can then get his grip on the difference between his changing his mind and a sensation's undergoing unnoticed change - in which case there is no phenomenological difference - by supposing that the same distinction applies in the unnoticed as in the noticed case. The extension from the seen to the unseen is no more problematic here than it is in the case of, say, the move from the observed to the unobserved present or from the known to the unknown past. Only a Dummettian anti-realism would threaten the coherence of this phenomenological point; and clearly such an anti-realism is threatening - if at all - to much more than the domain of unnoticed sensations. A satisfying riposte to it will meet

it on a broad front: it is quite unclear that a special argument to publicity will be required as part of the overall strategy in order to deal specifically with the case of unnoticed sensation-change.

20 The appearance of triviality imported by the word 'appropriate' will disappear when we spell out what sort of events count as appropriate causes of pain.

21 Again, I do not see, contra McDowell [3], that anything of epistemological significance turns on whether we allow pain-behaviour to cover all behaviour, including pretence, which looks like the expression of pain; and then say that pain-behaviour is a constitutive but defeasible criterion for pain (the special sense of 'criterion' commonly seen in Wittgenstein's use of the word): or whether (as McDowell prefers) we tie pain-behaviour strictly to actual pain - so that it is logically sufficient for the existence of a pain - and then locate the possibility of pretence in the possibility of mimicking pain-behaviour. The fact is that there is a non-contingent relation of some sort between some behaviour and pain - that is common ground between the models, as is the recognition that we can be deceived (so that a doubt about whether someone is in pain is always possible, if not always reasonable). There is no epistemological gain in taking McDowell's line, since a sceptic simply transfers his attack to the subject's 'knowledge' that he is confronted with pain-behaviour rather than imitation pain-behaviour. McDowell can only reply that if I am confronted with pain-behaviour (and I can only know this sort of

fact up to scepticism), then I am in epistemic touch with another's pain. But the position McDowell attacks has as much to say for itself. There I can know that I am confronted with pain-behaviour, and also know (up to scepticism) that it is expressive of pain, because it is embedded in suitable circumstances. The behaviour on this model does not function as an epistemological veil, precisely because of the overall constitutive connections between such behaviour and pain. If we take seriously this special sense of 'criterion', then I can be as closely connected with other minds on this picture as I can on McDowell's preferred picture.

22 As language seems to cut loose from the subjective features of experience: our language for colour is keyed to our discriminatory abilities and not to qualia. Cf chapter 3, note 40. That the subjective features of experience do not show up in language, at least not in this way (we would continue to call pillar-boxes 'red', even if we each saw different shades, provided we made the same discriminations), does not eo ipso impugn their ontological status, as Wittgenstein and his followers seem to think; see, e.g., Malcolm's contribution to Armstrong and Malcolm, at p.45ff.

23 Especially in Winch, Guttenplan and Wiggins [3], essay IV.

REFERENCES

- Albritton, R: 'On Wittgenstein's use of the term 'Criterion'' in
Pitcher.
- Armstrong, D and Malcolm, N: Consciousness and Causality (Oxford:
Blackwell, 1984).
- Ayer, A: [1] The Problem of Knowledge (Harmondsworth: Penguin,
1956).
[2] The Concept of a Person (London: Macmillan, 1963).
- Baker, G: 'Criteria: a new foundation for semantics' Ratio 1974.
- Berkeley, G: A New Theory of Vision etc. (London: Dent, 1910).
- Blackburn, S: Spreading The Word (Oxford: Clarendon, 1984.)
- Block, N: Readings in the Philosophy of Psychology vols 1 and 2
(London: Methuen, 1980).
- Brentano, F: Psychology from an Empirical Standpoint ed. L.McAlister
(London: Routledge, 1973).
- Campbell, K: 'Colours' in ed. Brown and Rollins Contemporary
Philosophy in Australia (London: Allen and Unwin,
1969).
- Castaneda, H-N: ''He': a Study in the Logic of Self-Consciousness'
Ratio 1966.
- Davidson, D: [1] Essays on Actions and Events (Oxford: Clarendon,
1982).
[2] Inquiries into Truth and Interpretation (Oxford:
Clarendon, 1984).
- Davies, M: Meaning, Quantification, Necessity (London:
Routledge, 1981).
- Denyer, N: Time, Action and Necessity: a Proof of Free Will (London:

- Duckworth, 1981).
- Descartes, R: [HR] The Philosophical Works of Descartes trans. Haldane and Ross vols 1 and 2 (Cambridge: CUP, 1911).
- [K] Philosophical Letters ed. A. Kenny (Minneapolis: University of Minnesota Press, 1981.)
- Doney, W: ed Descartes: A Collection of Critical Essays (New York: Doubleday, 1967).
- Dummett, M: [1] Truth and Other Enigmas (London: Duckworth, 1978).
- [2] 'Common Sense and Physics' in MacDonald.
- [3] 'Causal Loops' in Flood and Lockwood.
- Evans, G: [1] 'Things Without the Mind' in Van Straaten.
- [2] The Varieties of Reference (Oxford: Clarendon, 1982).
- Evans, G and McDowell, J: eds Truth and Meaning (Oxford: Clarendon, 1976).
- Flew, A and Vesey, G: Agency and Necessity (Oxford: Blackwell, 1987).
- Flood, R and Lockwood, M: eds The Nature of Time (Oxford: Blackwell, 1986).
- Frege, G: 'Der Gedanke' in Logische Untersuchungen transl. Geach and Stoothoff, (Oxford: Blackwell, 1977).
- Gale, R: The Language of Time (London: Routledge, 1968).
- Goldman, A: 'Perceptual Objects' Synthese 35 1977 pp.257-284.
- Grice, H: 'The Causal Theory of Perception' reprinted in Warnock.
- Grünbaum, A: 'The Meaning of Time' in Freeman and Sellars eds Basic Issues in the Philosophy of Time (Illinois: Open Court, 1971).
- Guttenplan, S: 'Moral Realism and Moral Dilemmas' PAS 1979-80.

- Harrison, B: Form and Content (Oxford: Blackwell, 1973).
- Healey, R: 'Quantum Mec^hanics and the Directedness of Time' in Healey
ed. Reduction, Time and Reality
(Cambridge: CUP, 1981).
- Hinton, J: Experiences (Oxford: Clarendon, 1973).
- Horwich, P: Asymmetries in Time (Massachusetts: MIT Press, 1987).
- Kant, I: Critique of Pure Reason (transl. Kemp Smith, London:
Macmillan, 1929).
Prolegomena to any Future Metaphysic transl. Carus,
Ellington (Indianapolis: Hackett, 1977).
Critique of Practical Reason (transl. Beck, New York:
Macmillan, 1985)
- Kaplan, D: 'Demonstratives', unpublished.
- Kripke, S: [1] Naming and Necessity (Oxford: Blackwell, 1980)
[2] Wittgenstein on Rules and Private Language (Oxford:
Blackwell, 1982)
- Lewis, D: 'Counterfactual Dependence and Time's Arrow' Nous vol.
XIII no.4 1979.
- Locke, J: An Essay Concerning Human Understanding ed. A.D.Wozley
(London: Fontana, 1964).
- Lucas, J: 'The Open Future' in Flood and Lockwood.
- MacDonald, G: Perception and Identity (London: Macmillan, 1979).
- McDowell, J: [1] 'On the Sense and Reference of a Proper Name' Mind
1977.
[2] 'Truth-Value Gaps' in Los and Pfeiffer eds Logic,
Methodology and Philosophy of Science, 6
(North Holland, 1981).

- [3] 'Criteria, Defeasibility and Knowledge' Proceedings of the British Academy 1982.
- [4] 'Values and Secondary Qualities' in Honderich ed. Morality and Objectivity (London: Routledge, 1985).
- [5] 'Singular Thoughts and Inner Space' in Pettit and McDowell.
- McGinn, C: [1] The Character of Mind (Oxford: OUP, 1982).
 [2] The Subjective View (Oxford: OUP, 1983).
- Mackie, J: [1] 'What's Really Wrong With Phenomenalism?' Proceedings of the British Academy 1969.
 [2] Problems From Locke (Oxford: Clarendon, 1976).
 [3] The Cement of the Universe (Oxford: Clarendon, 1980).
 [4] Logic and Knowledge: Selected Papers Vol.1 (Oxford: Clarendon, 1985).
- Malcolm, N: 'Descartes' Proof that his Essence is Thinking' in Doney.
- Martin, C and Deutscher, M: 'Remembering' PRev, 75, 1966.
- Maxwell, N: 'Can there be Necessary Connections between Successive Events?' in Swinburne ed. The Justification of Induction (Oxford: OUP, 1974).
- Mellor, D: Real Time (Cambridge: CUP, 1981).
- Melville, H: Billy Budd, Sailor and Other Stories (Harmondsworth: Penguin, 1967).
- Mill, J: An Examination of Sir Wm. Hamilton's Philosophy 6th ed. (London: Longmans, 1889).
- Millikan, R: Language, Thought and Other Biological Categories

(Cambridge, Mass: MIT Press, 1984).

- Nagel, T: The View From Nowhere (Oxford: OUP, 1986)
- Nozick, R: Philosophical Explanations (Oxford: Clarendon, 1981).
- Parfit, D: Reasons and Persons (Oxford: OUP, 1986).
- Pears, D: [1] Questions in the Philosophy of Mind (London: Duckworth, 1975).
- [2] The False Prison (Oxford: Clarendon, 1987).
- Penelhum, T: Survival and Disembodied Existence (London: Routledge, 1970).
- Perry, J: 'The Problem of the Essential Indexical' Nous vol.XIII no.1 1979.
- Pettit, P and McDowell, J: eds. Subject, Thought and Context (Oxford: Clarendon, 1986).
- Pitcher, G: ed. Wittgenstein, a Collection of Critical Essays (London: Macmillan, 1966).
- Popper, K: The Logic Of Scientific Discovery (London: Hutchinson, 1959).
- Prior, A: [1] Papers on Time and Tense (Oxford: Clarendon, 1968).
- [2] Objects of Thought (Oxford: Clarendon, 1971).
- [3] Papers in Logic and Ethics (London: Duckworth, 1976).
- Putnam, H: Reason, Truth and History (Cambridge: CUP, 1981).
- Quinton, A: [1] 'The Problem of Perception' repr. in Warnock.
- [2] The Nature of Things (London: Routledge, 1973).
- Russell, B: [1] The Problems of Philosophy (Oxford: OUP, 1967).
- [2] Mysticism and Logic (London: Unwin, 1986).
- Schopenhauer, A: On the Freedom of the Will transl. Kolenda (Oxford: Blackwell, 1985).

- Sellars, W: [1] 'Time and the World Order' in ed. Feigl and Maxwell,
Minnesota Studies in the Philosophy of Science
vol.III (Minneapolis: University of Minnesota
Press, 1962).
- [2] Science, Perception and Reality (London: Routledge,
1963).
- Shoemaker, S: 'Functionalism and Qualia' in Block vol.1.
- Shoemaker, S and Swinburne, R: Personal Identity (Oxford: Blackwell,
1984).
- Sprigge, T: 'Personal and Impersonal Identity' Mind 1988.
- Stevenson, L: The Metaphysics of Experience (Oxford: Clarendon,
1982).
- Strawson, G: Freedom and Belief (Oxford: Clarendon, 1986).
- Strawson, P: [1] Individuals (London: Methuen, 1959).
[2] The Bounds of Sense (London: Methuen, 1966).
[3] Freedom and Resentment (London: Methuen, 1974).
[4] 'Perception and its Objects' in MacDonald.
[5] 'Reply to Evans' in Van Straaten.
[6] 'Causation and Explanation' in Vermazen and
Hintikka eds. Essays on Davidson:
Actions and Events (Oxford: Clarendon, 1985).
- Van Straaten, Z: ed. Philosophical Subjects (Oxford: Clarendon,
1980).
- Vesey, G: ed. Knowledge and Necessity (London: Macmillan, 1970).
- Walker, R: Kant (London: Routledge, 1978).
- Warnock, G: ed. The Philosophy of Perception (Oxford: OUP, 1967).
- Whitrow, G: The Natural Philosophy of Time (London: Nelson, 1961).

- Wiggins, D: [1] 'The De Re Must: a note on the logical form of Essentialist Claims', in Evans and McDowell.
- [2] Sameness and Substance (Oxford: Blackwell, 1980).
- [3] Needs, Values, Truth (Oxford: Blackwell, 1987).
- [4] 'Freedom, Knowledge, Belief and Causation' in Vesey.
- Wilkes, K: Physicalism (London: Routledge, 1978).
- Williams, B: [1] Problems of the Self (Cambridge: CUP, 1973).
- [2] Descartes (Harmondsworth: Penguin, 1978).
- Williams, M: Groundless Belief (Oxford: Blackwell, 1977).
- Winch, P: 'The Universalizability of Moral Judgments' Monist 1965.
- Wittgenstein, L: TLP = Tractatus Logico-Philosophicus (London: Routledge, 1922).
- PI = Philosophical Investigations transl. Anscombe (Oxford: Blackwell, 1958).
- OC = On Certainty (Oxford: Blackwell, 1979).
- Z = Zettel (Oxford: Blackwell, 1981).
- Wollheim, R: The Sheep and the Ceremony (Cambridge: CUP, 1979).
- Wright, C: [1] 'Anti-realist semantics: the role of Criteria' in Idealism: Past and Present ed. Vesey (Cambridge: CUP, 1982).
- [2] 'Facts and Certainty' Proceedings of the British Academy 1985.
- [3] 'A Cogent Argument against Private Language?' in Pettit and McDowell.

