Mind as Metaphor: A Physicalistic Approach to the Problem of Consciousness

Jim Hopkins

In what follows I present an approach to the problem of consciousness, which I take to be suggested by Wittgenstein's remarks on sensation. As sketched here, this consists of a number of empirical hypotheses about the mind and how we represent it, and a series of arguments that these hypotheses explain phenomena which constitute the problem of consciousness, in such a way as to render them neither mysterious nor problematic.

If hypotheses and arguments of this kind were sufficiently complete and correct they should constitute a resolution of the problem. As might be expected, those here fall short of this goal. Among other things they are incomplete, vague, speculative, and likely to contain a number of errors. Nonetheless they seem to me worth putting forward as well as I can. I think that an approach of this kind has greater explanatory potential than has been appreciated, and that even if the present proposals are mistaken they may still provide some indication as to how the problem might be better treated.1

I

The problem of consciousness arises from the nature of experience itself. Experience presents us with a world which we take to be *objective* and *physical*. The experiences which present this world, by constrast, seem *subjective* and *phenomenal*. In having experience we take ourselves to encounter purely subjective aspects or qualities -- the way pain feels to us, the way colour and shape look to us, the way things sound to us, and so forth -- which are *sui generis*, and utterly different from the physical things or processes which they represent. Hence the origin of such *qualia* in the brain, or their relation to the physical world in general, seems incapable of explanation.

This radical difference between experience and the physical world, and our consequent sense of the inexplicability of experience, have long been evoked and discussed in philosophy. Thus in his *New Essays on the Understanding* of 1704 Leibniz urged that thinking or sensation

...cannot be an intelligible modification of matter, or one which could be understood and explained; that is to say, a sentient or thinking being is not a mechanical thing like a watch or a windmill, so that we could conceive of sizes, shapes, and motions in such a mechanical conjunction that they could produce...thinking and sensing [which] would likewise stop if the mechanism got out of order. Thus it is not natural to matter to have sensation or to think...2

And in the *Monadology* he made the point more vivid via a celebrated thought-experiment.

...Suppose that there were a machine so constructed as to produce thought, feeling, and perception, we could imagine it increased in size while retaining the same proportions, so that one could enter as one might a mill. On going inside we should only see the parts impinging upon one another; we should not see anything which would explain a perception...3

It is no defect in Leibniz's argument that the brain does not work like a mill, for his conclusion apparently applies quite generally. We seem to have no conception as to how any physical alteration or process might serve to 'explain a perception.' For this reason progress in understanding the brain has been accompanied by a series of more modern versions of Leibniz's claim. In 1866 the physiologist Julian Huxley observed that

...How it is that anything so remarkable as a state of consciousness comes about as a result of irritating nervous tissue, is just as unaccountable as the appearance of Djin when Alladin rubbed his lamp.4

In 1872 another eminent physiologist, Emile Du Bois-Reymond, put the idea in a slightly fuller context, urging that the evolution of conscious life constitutes a point at which

...our knowledge of nature reaches a gap to be crossed by no bridge, no wing...consciousness cannot be explained from its material conditions, not only -- as everyone will admit -- at the present state of our knowledge, but according to the very nature of things.

In support of this he asks

What conceivable connection is there between certain movements of certain atoms in my brain on one side, and on the other the original, undefinable, undeniable facts: 'I feel pain, feel lust; I taste sweetness; smell the scent of roses, hear the sound of an organ, see redness'?

and answers

...It is entirely and forever incomprehensible why it should make a difference how...atoms are arranged and move, how they will be arranged and move. It is in no way intelligible how consciousness may arise from their coexistence.5

And such remarks may be compared with a recent statement by Christof Koch.

...the really difficult aspects [of consciousness], like subjective feelings, may not have a scientific solution. The subjective state...of pain, of pleasure, of seeing blue, of smelling a rose -- there seems to be a huge jump between the materialistic level, of explaining molecules and neurons, and the the subjective level. Let's focus on things that are easier to study...6

Similar sentiments have been propounded by countless other investigators. But in recent years, as the case for the view that conscious events are neural events has been strengthened by both philosophical argument and neuroscientific progress, Leibniz's claim has been repeated with increasing insistence and frequency. Thus in his celebrated article 'What is it Like to be a Bat?' Thomas Nagel stressed that 'the subjective character of experience' is 'not captured by any of the familiar, recently devised reductive analyses of the mental'; 'not analyzable in terms of any system of functional states, or intentional states'; and 'not analyzable in terms of the causal role of experiences'. Hence, as he urges, 'If we acknowledge that a physical theory of mind must account for the subjective character of experience we must admit that no presently available conception gives us a clue as to how this could be done.'7 Likewise Colin McGinn asks 'How can technicolour phenomenology arise from soggy grey matter? Somehow, we feel, the water of the physical brain is turned into the wine of consciousness, but we draw a total blank on the nature of this conversion. Neural transmissions just seem the wrong kind of matrial with which to bring consciousness into the world...'8 Or finally as David Chalmers writes in a recent issue of the *Scientific American*

From an objective viewpoint, the brain is relatively comprehensible. When you look at this page, there is a whir of processing: photons strike your retina, electrical signals are passed up your optic nerve and between different areas of your brain, and eventually you might respond with a smile, a perplexed frown or a remark. But there is also a subjective aspect. When you look at the page you are conscious of it, directly experiencing the images and words as part of your private, mental life. You have vivid experiences of colored flowers and vibrant sky....The hard problem...is the question of how physical processes in the brain give rise to subjective experience. This puzzle involves the inner aspect of thought and perception...It is these phenomena which constitute the real mystery of the mind.9

This is the problem of consciousness. We have good reason to hold that experience is both caused and realized by neural processes in the brain. These processes are both physical in themselves, and the effects of encounters with the public, physical objects of experience. Yet experience itself seems constituted by phenomena which are neither public nor physical, but rather subjective, inner, and private. Hence consciousness seems inexplicable by reference to the brain, or indeed by any sort of physical account which we can envisage.

Should we accept that this impass is, as Du Bois-Reymond says, 'according to the very nature of things', so that consciousness cannot be explained at all? In the article mentioned above Nagel argued that creatures very different from us would possess very different kinds of consciousness, which we would be incapable of understanding. As he held

...there are facts which humans never will possess the requisite concepts to represent or comprehend...facts which could not ever be represented or comprehended by human beings, even if the species lasted forever -- simply because our structure does not permit us to operate with concepts of the requisite type.

McGinn has argued that we should apply this conception to the problem of consciousness itself. Like Du Bois-Reymond, McGinn takes the capacity of the brain to produce consciousness to be a natural

phenomenon which has arisen in evolution and is thus fixed in our genes. But so also, McGinn stresses, is the capacity for scientific understanding; and like any other product of evolution this capacity has natural limits. Understanding the production of consciousness, McGinn argues, is simply beyond these limits.

Other philosophers have been less pessimistic. Thus Jerry Fodor as well as Nagel has stressed the depth of the problem10, but both have also remarked that McGinn's conclusion may be premature, and that the problem may yield to some far-reaching and as yet unforseen change in our scientific understanding. This is a view with which a number of eminent scientists have expressed their agreement; and some have offered speculations as to the kinds of radical new hypotheses which might be required. Hence, as Thomas Metzinger writes in a recent anthology, the problem now seems to many to be 'the last great puzzle and the greatest theoretical challenge of our time,' requiring a new science to foster a new approach to mind and nature alike.11

...Many hold that the necessary revolution can only occur when our understanding of the subject transcends disciplinary boundaries and links between the relevant areas of research are drastically increased. However it has also become obvious that a systematic integration of research activities is necessary as well. This situation has led to a call for a new science, *the science of consciousness*, both from empirical and theoretical researchers.12

II

The problem of consciousness -- the apparent physical inexplicability of the phenomenal properties of experience, as articulated by Leibinz and very many others -- thus seems both clear and well recognized. The sense of inexplicability which constitutes the problem seems rooted in an intuition which is widespread, longstanding and powerful. This is simply that there is a very striking distinctness or heterogeneity -- a notable and radical difference in kind -- as between the subjective and phenomenal nature of experience, and the objective and physical nature of the world. (Thus for example we have Leibniz's notion that sensation 'is not natural' to matter, so that the former cannot be 'an intelligible modification' of the latter; du Bois-Reymond's 'gap to be crossed by no bridge' between 'the original, undefinable, undeniable facts' of subjective experience and its 'material conditions' in the brain; Koch's 'huge jump' between the 'subjective' and 'materialistic' levels; McGinn's opposition of the 'technicolour phenomenology' of the 'wine of consciousness' to the 'soggy grey matter' of the brain, which seems 'the wrong kind of materials' for the production of consciousness; Chalmers' opposition of the 'subjective aspect of experience', involving 'vivid experience of colored flowers and vibrant sky', to the 'relatively comprehensible' physical brain; and so on.) The subjective character of experience seems to us evident, vivid and significant; we naturally differentiate this from the physical chatacteristics of the world in general and the brain in particular; and so we arrive at our sense of a gap between consciousness and the physical, which explanation cannot bridge.

Let us call this the *intuition of difference* as between phenomenological (subjective, introspectible) and

physical things or properties. We can appreciate the constitutive role of this intuition if we consider Nagel's remark that 'If mental processes are physical processes, then there is something it is like, intrinsically, to undergo certain physical processes. What it is for such a thing to be the case remains a mystery.' To a physicalist this ought to seem a puzzling way of putting the matter; for according to physicalism nothing could be more natural, or more to be expected, than that there is something it is like, intrinsically, to undergo certain physical processes. If pain is realized by neural events of kind P, then there must be something it is like to undergo an event of kind P, for this is to feel pain. Yet even if we identify pains with events of kind P, we still encounter the epistemic 'gap to be crossed by no bridge'. We feel that knowing all about the neural kind P would still not provide us with an explanation of how pain feels, or, to put the point as Frank Jackson does, that someone who had never felt pain could learn all about the neural kind P and yet still not understand what pain felt like, that is, still not know the phenomenological quality of pain.13 And this is because we feel that the phenomenological quality of pain -- or any other introspectible property -- is radically different from any physical kind, whatever that kind happens to be.

It seems clear both that very many people have this intuition, and that in response to it we are inclined to make two distinctions. The first is almost immediate, and the second follows as a result of further thought. First, since we take the subjective nature of experience to be radically distinct from the physical nature of the worldly objects of experience, we construe experience as constituting a realm of inner mental representations, which are distinct from the physical world they represent. Then secondly, as we reflect upon the differences between these representations and the things they represent, we seek to distinguish two kinds of properties which we represent things as having. We try to separate the *real* or *primary* properties, which characterize the physical nature of things as they are apart from our mental representations of them, from other merely *apparent* or *secondary* properties, which do not characterize the intrinsic physical nature of things, but are rather somehow derived from the way things are presented to us via the introspectable character of experience. (To put the idea in terms of a metaphor: Once we accept that worldly things are presented to us via phenomenal spectacles, we seek to distinguish the real properties of these things from the apparent properties deriving from the stained glass which mediates our apprehension.)

The force of this intuition, and the naturalness of the distinctions which follow in its wake, are attested by the pervasive influence they have exerted over our thought about mind and the world. In almost every era philosophers or scientists who have reflected upon the physical causation of experience have tended to distinguish between the phenomenal character of experience and the physical world, and hence also to distinguish the properties which objects really possess from those which are artefacts of their phenomenal representation. We find such distinctions in ancients such as Democritus and Epicurus; medievals such as Augustine and Acquinas; moderns such as Galileo, Hobbes, Descartes, Boyle, and Locke; and thinkers of the present day. Here, for example, is a recent discussion by two eminent evolutionary psychologists:

Far from being a physical property of objects, color is a mental property -- a useful invention that specialized circuitry computes in our minds and "projects onto" our percepts of physically colorless objects...What is true for colour is true for everything in our experienced worlds: the warmth of a smile, the meaning of a glance, the heft of a book, the force of a glare...We inhabit mental worlds populated by

the computational outputs of battalions of evolved, specialized, neural automata. They segment words out of a continual auditory flow, they construct a world of local objects from edges and gradients in our two-dimensional retinal arrays...Oblivious to their existence, we mistake the representations they construct (the color of a leaf, the irony in a tone of voice, the approval of our friends, and so on) for the world itself...14

In this passage the sense of difference between phenomenal and physical is re-expressed in the conception of a related oppositon between two kinds of properties (mental and physical) of things in the world.15 While Chalmers takes the perception of vivid colour to be part of the subjective, inner, private aspect of experience, Cosmides and Tooby emphasize an apparent consequence of such a view, namely that colour (or at least the vivid subjective phenomenal colour to which Chalmers refers) is not an aspect of the objective, outer, public world. As Chalmers holds that we 'directly' experience colour as part of a private, inner, mental life, so Cosmides and Tooby hold that we experience colour via its projection into 'percepts' or 'representations', which constitute 'experienced worlds' or 'mental worlds' distinct from 'the world itself'. (Hence on these accounts we are mistaken if we suppose that colour, or the quality of colour as subjectively apprehended, is a feature of the world.) Cosmides and Tooby thus postulate and describe a physiological and psychological process -- the production and projection of colour by the battalions of the brain -- of which they can give no further account. For this, evidently, is another version the original problem of consciousness, that is, the question as to how 'anything so remarkable as a state of consciousness' ('the wine of consciousness', the 'inner aspect of thought and feeling', etc.) results from processes in the brain.

Authors have expressed the intuition of difference, and the distinctions which attend it, in many ways. (Perhaps the briefest is Berkeley's 'Nothing but an idea can be like an idea'.) In one form or another, however, the intuition seems to have shaped almost all accounts of experience, and so to have constantly rendered problematic the relations between the mental and the physical, and mental representation and worldly reality.16

III

Wittgenstein seems to have been the first philosopher both to acknowledge the intuition of difference and also to submit it to extended critical examination. Thus in *Philosophical Investigations*17 he writes:

412. The feeling of an unbridgeable gulf between consciousness and brain-process: how does it come about that this does not come into the considerations of our ordinary life? This idea of a difference in kind is accompanied by a slight giddiness, -- which occurs when we are performing a piece of logical sleight-of-hand....

The claim here is that the intuition -- the feeling of 'a difference in kind' which constitutes 'an unbridgeable gulf' as between consciousness and brain process -- is not to be taken at face value, but

rather to be regarded as involving some kind of confusion or mistake, a 'piece of logical sleight-of-hand'. Hence Wittgenstein (and some who have followed his lead, such as Dennett and Papineau18) regard the problem of consciousness quite differently from the philosophers and scientists mentioned above. Not as 'the greatest theoretical challenge of our time', but rather as a form of confusion, rooted ultimately in misunderstanding of our own language or thought. This represents wrestling with the problem as a less glamorous form of activity, and has been unpopular among students of consciousness, particularly those aspiring to the vanguard of the revolution.

Wittgenstein's arguments on this head can be divided into two parts, which we will consider briefly later. On the one hand, he attempts to show that the intuition has consequences which render it unacceptable and self-defeating; and on the other he attempts to characterize it as arising from a way of representing the mind -- a metaphor or 'picture' of thought, feeling, and so forth -- which cannot literally be applied.19 Thus he says:

427. "While I was speaking to him I did not know what was going on in his head." In saying this one is not thinking of brain-processes, but of thought-processes. The picture should be taken seriously. We should really like to see into his head. And yet we only mean what elsewhere we should mean by saying we should like to know what he is thinking. I want to say: we have this vivid picture -- and that use, apparently contradicting the picture, which expresses the psychical.

He spells out the nature of the 'picture' he takes to be involved here more fully in the course of criticizing its role in our thinking about the mind.

293. If I say of myself that it is only from my own case that I know what the word "pain" means -- must I not say the same of other people too? And how can I generalise the one case so irresponsibly?
Now someone tells me that *he* knows what pain is only from his own case! --- Suppose everyone had a box with something in it: we call it a "beetle". No one can look into anyone else's box, and everyone says he knows what a beetle is only by looking at his beetle. -- Here it would be quite possible for everyone to have something different in his box. One might even imagine such a thing constantly changing. -- But suppose the word "beetle" had a use in these people's language? -- If so it would not be used as the name of a thing. The thing in the box has no place in the language-game at all; not even as a *something*: For the box might even be empty. -- No, one can 'divide through' by the thing in the box; it cancels out, whatever it is.
That is to say: if we construe the grammar of the expression of sensation on the model of 'object and designation' the object drops out of consideration as irrelevant.

The 'picture' Wittgenstein here criticizes is that of the mind as an *enclosed space* or *container*, whose contents are in themselves phenomenological and psychological rather than physical ('one is not thinking of brain-processes, but of thought-processes'), and detected by a process analogous to sight. In this representation each of us can introspect the psychological contents of his or her own inner space or container, but not those of another, so that the contents of this space or container are private. It does seem that we tend to picture the mind in some such way. This is shown, for example, in the way readers have constantly interpreted Wittgenstein's remark as describing what the mind is actually like -- as bringing out

the way in which each of us can in fact know pain only in our own case, and the consequences of this fact for language -- rather than as a *reductio ad absurdam* of the view.

To say that we use this picture is not to say that we suppose it literally true. Most, for example, hold that we do not see phenomenal qualities, but rather *introspect* them, so that the process of detection is distinguished from activity involving the eyes. But the term 'introspection' derives from the Latin for seeing into, and we constantly compare it to vision. Thus McGinn speaks of phenomenology as 'technicolour'; and Chalmers uses both visual and spatial metaphors, emphasizing 'the inner aspects of thought and feeling,' and urging that '...information processing...does not go on in the dark. There is also an internal aspect...This internal aspect is conscious experience.'20 If we ask what 'inner' or 'internal' mean here, the answer seems to be that these aspects of thought and feeling are internal to -- that is, in some sense inside -- the space or container which is the mind. (Again, the 'experienced worlds' or 'mental worlds', which Cosmides and Tooby take us to 'inhabit', seem also to be conceived in spatial terms, i.e. as somehow surrounding us, while not being the real 'world itself.') It seems that we very often think and speak in these ways, but without giving fully literal credence to the way of thinking implied. Thus consider the use of the concept of space in the following from Metzinger:

To be able to speak seriously about a *science of consciousness*, a number of fundamental questions would have to be answered. It is interesting to note that with the emergence of consciousness private worlds -- spaces of inner experiences -- are opened up. These spaces, however, are *individual spaces*: ego-centres of experience that suddenly appear in a centerless universe. Each such centre of consciousness constitutes its own perspective on the world. This perspective is what philosophers sometimes like to call the 'first-person perspective'. A phenomenal world of its own is tied to each of these perspectives. These individual worlds of experience also possess a historical dimension: almost always a psychological biography emerges together with them -- what we call our 'inner life'. This too can be seen as the history of the genesis of a world, or a *phenomenal* cosmology: within each of us a cosmos of consciousness unfolds temporarily, a *subjective* universe develops. The first part of the problem is to understand how a variety of subjective universes can constantly form and disappear in our objective universe...21

It thus seems, as Wittgenstein held, that we make frequent use of metaphors or modes of comparison in which we liken the mind to a space or container within which phenomenological or subjective properties are located.22 To better appreciate the importance of this it will be useful now to consider some work on metaphor and thought.

IV

A number of cognitive scientists, including George Lakoff, Mark Turner, and their colleagues, have recently argued that metaphor should not be seen as a linguistic device, but rather as a form of thought, which is pervasive, systematic, and fundamental.23 Their claim is that we frequently think about objects, properties, or relations in one domain (called the *target domain*) by systematically mapping these onto

objects and properties in another domain (called the *source domain*). The correspondence relation between these domains constitutes a potentially large and organized *conceptual metaphor*, by means of which we think, or conceive, the one domain in terms of the other.

Where the source domain is **A** and the target **B**, so that in mapping the domains we think of **B** in terms of **A**, we can speak of the **B as A** metaphor. Thus, to take one of Lakoff's examples, we appear to make use of a metaphor of love-as-journey. In this use concepts of objects, properties, and relations from the domain of *travel* or *journeys* in order to conceptualize objects, properties, and relations in the domain of *co-operative personal relationships*, including in particular relations of *love*. In this we systematically take *persons* in such relations to correspond to *travellers*, their *relationship* to the *vehicle* in which they are travelling, and their goals in the relationship to their *destinations* in travelling. Thus we may speak of such a relationship as *going along well*, *slowing down*, *going nowhere*, *getting stuck*, *at a crossroads*, *at a dead end*, and so on.

We also reason in accord with this mapping. We take it, for example, that if a relationship is *stuck*, those involved have reason to try to do something about this. They may try *start over* or to *get the relationship started again*, or to *get going* or *going forward*, once more. *Towards this end* they may, for example, try to get over the problem, or to find their way out of the difficulty. Alternatively they may decide that the relationship has *broken down*, or perhaps been *wrecked* by the actions of one or both of them, in which case they will *get out* and *go their separate ways*.

In this metaphor, as in many others, the source domain in terms of which we think is intuitively more concrete than the target domain which we think about. Also we relate the domains tacitly, in the sense that we may be unaware both of using such correspondence relations, and of their richness and systematicity. Thus we may tacitly represent a relationship by one or another sort of vehicle, as seems appropriate to the rest of out thought. The relationship may be *taking off* (airplane); *on the rocks* (boat); *off the rails* (train); *in the slow lane* (car); and so on. Further, as our understanding of words enables us to form and understand new sentences, so our understanding of the appropriate correspondence relations enables us to form and understand new instances of metaphor. Thus, to take another of Lakoff's examples, in hearing a song lyric like 'We're driving in the fast lane on the freeway of love', we are immediately able to understand the metaphor in terms of the relationship-as-journey structure, and others associated with it. The vehicle (relationship) is going fast, and this is connected with excitement (fast cars, fast women); this speed is compared to fast or reckless driving, which may lead to a crash in which someone will be hurt; the idea that the road is a freeway links with the idea of sexual freedom, free love, and so on. The tacit grasp of underlying correspondence relations thus enable us to understand unfamiliar instances of a conceptual metaphor, and without explicit awareness of the sources or details of our understanding.

The hypothesis that we think in terms of correspondence relations of this kind enables us to explain a variety of linguistic and conceptual data, including our systematic use of linguistic expressions and patterns of inference.24 Also when we delineate a given metaphor on the basis of certain data, we can usually find further data which the postulation of that metaphor serves to explain. Thus for example we can ask why the phrase 'spaceship earth' has the particular resonance it does. Part of the answer seems to be that this phrase evokes the idea that human beings have an important co-operative relationship in virtue

of being inhabitants of a common planet. This evocation is explicable on the hypothesis that the phrase constitutes an instance of the correspondence between relationship and vehicle in the metaphorical structure we have been discussing. (Compare the more explicit 'We're all in the same boat.') Again, one might ask why the song which begins 'Trains, and boats, and planes, all bound for Paris, New York, and Rome...' should be so evocative of lonliness or solitude. And again there seems a plausible answer, in terms of the same metaphor: these vehicles which the singer is not in, going on journeys which the singer is not taking, represent relationships in which the singer, sadly, has no part. A few such examples may not seem particularly convincing; but each structure of the kind we are considering seems to generate an open class of examples, which have considerable cumulative weight.

The metaphor we have just been discussing -- of love, or relationship more generally, as journey -- seems to occupy a particular place in a system of metaphors involving movement. This metaphor has a number of close relations, such as that of career-as-journey. In this metaphor we use of a number of expressions (*climbing* in one's occupation, getting to a *higher* level, getting *to the top*) which are also aspects of another pervasive metaphor, that of better-as-higher (high status, high profile, high achiever, etc.). These metaphors, in turn, can be seen as part of a more general correspondence relation, that in which long-term purposeful activities are compared with journeys. (He had raised a lot of money, but he still had *a long way to go* towards raising the sum he needed.)

A particularly inclusive instance of this metaphor is that of a purposeful life as a journey. In this we depict life as a journey through various locations, and represent various goals as destinations which we can succeed or fail to reach in various ways. Thus Dante begins the *Inferno* by saying 'In the middle of the journey of our life I came to myself in a dark wood where the straight way was lost....'. In reading this we know that he is speaking of his purposes or goals in life. Again, when Robert Frost writes that

```
Two roads diverged in a wood, and I -
I took the one less travelled by,
And that has made all the difference.
```

we can see that he is talking about a time of decision in life, which was resolved by the choice of a way of life -- presumably that of a poet -- which was less conventional ('less travelled by'). (Frost also uses the metaphor of the wood, as a place in the journey of life in which the question of keeping to the right way or path becomes salient; and he apparently indicates both how important the choice was for his self, and his hesitation in making it, by a repetition of 'I'. The reader thus partly enacts the hesitation, decision, and affirmation of self which the poet is describing, by hesitating and then going ahead in reading the instances of 'I' in the first and second lines.)25

This metaphor of life as journey can in turn be seen as an instance of a more inclusive correspondence, in which we represent goals as destinations of one kind or another, towards which agents move. In this metaphor we represent purposive activities as one or another sort of movement towards goal-destinations, means of achieving goals as paths or ways to these destinations, difficulties in achieving goals as

impediments to motion, and so on.26 Thus difficulties in achieving goals are represented as *barriers*, which we may *run into*, *come up against*, and so forth, and which we may try to *get over* or *get around*, but by which we may be *stopped*, *impeded*, *blocked*, *held back*, *trapped*, *fenced in*, *boxed in*, or otherwise rendered unable to *get through* or *get ahead*. Also we may risk getting *tied up* with other tasks, which can also *side-track* us. We may also be *hindered* in our *progress* by features of the metaphorical terrain through which we are moving, for example if we encounter *hard going* or *rough going*, or if things happen to be *uphill*. Again we may be *loaded down* with *burdens* of various kinds, such as duties which *weigh on us*, or again we may have obligations which *tie us down* or problems which *hang us up*. Again, as Dante and Frost point out, in any *pursuit* we may lose our way, or fail to *find the right path*, or be *blown off course*, or whatever.

This metaphor -- purposeful activity as movement towards goal -- is again highly organized. Within it manner of action corresponds to manner of motion, different means for action correspond to different paths, things which affect action are things which affect motion, the inability to act is an inability to move, and so forth. Each of these consequences shows up in detailed metaphoric usage. Thus for example making *progress* is forward movement, e.g. *moving ahead* or *forging ahead*, and the amount of progress is the distance moved, as when one has *come a long way*, or *covered a lot of ground*, in some project. Accordingly failure to make progress is failure to move, as when things are at a *standstill*, or *stuck*, or *going nowhere*. We are particularly likely to get this state is someone gives us false information pertaining to pursuing our goals, and so *misdirects* or *misleads* us, so that we thing we are making progress when we are not, and may indeed be *going in the wrong direction*, or just *going around in circles*. Likewise lack of purpose is lack of *direction*, as when one is just *drifting aimlessly*. Again, beginning a project is *starting out*, or *taking the first steps*, and finishing it is *reaching the end* which at the beginning may have seemed a *long way off*. A contrary of progress is *backsliding*, or any form of *going in the wrong direction*, which may mean that we have to *retrace our steps*, or *go back and start again*.

Another aspect of this metaphor seems to be the mapping of states one may be *in* with places which one may enter, leave, etc., in one's purposive activity/movement. Thus one may *get through* a depression, that is *emerge from* the depressed state after one has *entered* it. This is also a form of progress construed as movement in the direction of a goal. Again, we think of a bad mood as something one may be *bogged down* or *stuck* in, and depression, like love, as something one may *fall into* and not *get out of* again. (In the case of love the state is a valued one, so that movement out of it does not count as progress unless this is actually a desired goal.)

This range of metaphors is important for our topic because it corresponds with the basic fact, central to the philosophy of action, that we achieve a vast range of our goals by moving our bodies in particular ways (cf Davidson's remark in 'Agency' that 'We never do more than move our bodies; the rest is up to nature.'27). Through this meatphor we gain focus on the central role of the human body, and various actions or movements of the human body, as basic source domains in the system of metaphorical thinking which we are considering.28

These metaphors for purposive activity represent the world as from the perspective of an agent engaged in

motion towards things in the world. This perspective can be contrasted with another, in which the agent is *standing still* or *stationary* in relation to things. In the perspective of the agent as moving, our motion is not only towards the fulfilment of goals, but also towards future events generally, which we thereby represent as stationary objects. In the perspective of the agent as stationary, by contrast, we represent goals and events as objects which move towards us. (This contrast is sometimes also linked with that between activity and passivity.) In the first perspective, for example, we are *approaching Christmas*, or *moving towards* completion of a project; while in the second *Christmas is coming* or completion is on its way. In the first perspective, again, we are *nearly there*; whereas in the second, the goal, expected event, or whatever, is *nearly here*. (These perspectives can be combined, so that we represent events as objects which move in relation to us as we also move, for example when we are *going with the flow of events*, or again if *things are going against us*.) The contrast applies also to the representation of states as locations: in the perspective in which the agent is stationary, one does not *go through* a depression or comparable state; rather the state *comes on*, and after a time perhaps *goes away*.

A related perspective of body-involving metaphor is that of an agent who is close enough to things to take hold of them. In this perspective we are concerned with goals which are *at hand*, and so *within one's grasp*, so that what we have to do is to *reach out* and *seize* them. (To see how metaphorical this is, remember that one may *seize*, or *let slip*, a chance, an opportunity, some time, or the day; these are things which we may either *grab*, *grasp*, *get a grip on*, or *take hold of*, or which may *slip through our fingers*.) Since in grasping something one takes possession of it, this system of metaphor represents goals, or opportunities to achieve goals, as possessions (or potential possessions) which one may *get* or *have*, or again which may be *given to* one or *taken away* from one. Someone who *seizes* an opportunity to *get to the top* may thus be worried that a rival may *steal* or *occupy* the position he is seeking.

Again this metaphor applies to states: as long as going to a party doesn't *take* my bad mood *away*, I will still *have* it, whether I am trying to *hold on* to it or not. In the perspective in which the agent is moving we seek to *avoid* problems, but may *run into* them; but if problems are *at hand*, we may *get* them even if we don't want to. Thus we may say, in terms of location, that John wanted to *keep away* from trouble, and to *keep out* of it; but Tom *led* him *into* it, and once he was *in*, he had to rely on Joan to *help him out*. Alternatively, we may say in terms of possession that John didn't *want* (want *to get*, want *to have*) any trouble, but that Tom *gave* him a lot of it; and that once he was *having trouble*, Joan had to help him *get rid of it*.

The metaphorical perspective in which the agent *moves* in relation to goal- or state-locations is closely related to that in which the agent *grasps* and so *possesses* goals, states, or attributes. The two are connected via the double meaning, and hence the double metaphorical use, of 'reach'. This word seems to have acquired the meaning of arriving at a destination, and hence the metaphorical meaning of achieving a goal (*reaching home*, *reaching port*, *reaching the age of 21*, *reaching the end of the race*, or whatever) from the more primitive and bodily meaning of *reaching for* something with the hands, via a process of metaphorical extension related to the kind of thinking we are discussing.29 This illustrates what might be called a switching point in the system of bodily metaphors we are considering, where those related to the very basic bodily movements of manual reaching and grasping map to those related to the movements connected with walking, and hence with locomotion more generally.

It thus appears that our use of conceptual metaphor includes the mapping of general categories -- including those of time, event, state, property, and goal -- on to concrete and bodily activities of moving towards things, taking hold of them, and so forth. The body and its activities thus provide a source domain even for relatively abstract categories of thought. We relate this source domain via these general categories to various more specific target domains, and in this process we relate the target domains to more specific bodily activities, such as choosing paths, travelling in vehicles, climbing in heirarchies, and so forth. As we act by moving our bodies, so we think in terms of bodily movement as well.

V

The use of the body as source domain for conceptual metaphor clearly extends to the concept of mind. Thus we systematically relate intellectual achievements to perceptual achievements, and perceptual achievements to those of basic manual manipulation. We compare understanding, for example, to sight. In understanding we see what another means, and we are better able to do so if the other speaks or writes *clearly*. Understanding *illuminates* things for us, that is, it *casts light* on them; and this means that we can attain a sharper *perspective*, or a *better view*. In consequence we will not be *in a fog* or *haze* or *benighted* about things, as we imply that people were in the *dark* ages, that is, before the *enlightenment*.

The further importance of this in relation to the body is also *clear*, for when we can *see* something we can *get* it. *Seeing things clearly* enables us to *grasp them firmly*. Hence when someone *puts* things in a way which is *transparent* to us, we are thereby enabled to *get a hold* on the subject matter. We don't like it when people talk or reason in a *foggy* way, as this is liable to render things *obscure* to us, that is, to *cover* or *hide* the truth, and hence to leave us *in the dark*. Likewise if someone reasons in a *slippery* way, we will find it hard to *get a grip* on what he means. Such metaphorical relation of intellectual understanding to perceptual accuracy and bodily control seems to be present in many cultures and languages. This system also permeates the others which we have discussed: note for example Dante's statement that he *lost* his way in a *dark* wood. The same is true for other metaphors by which we think of the mind in terms of the body.

We considered Wittgenstein's example of the beetle in the box -- in which the mind is represented as a container, with the space inside the container thus represented as an *inner space* -- above. This metaphor seems part of a particularly significant family. A very familiar instance, for example, involves comparing the mind to *a house*. There is a joke in which we knock on the forehead of an inattentive or *vacant* person, asking if anyone is *at home*. (There is also a children's game along the same lines, which even very little children instantly understand and enjoy.) Again, when a person's mind is not present, in one way or another, we may say that *the lights are on* (the eyes are open) *but there is no one at home*. Likewise we speak of the *house* of reason; of the mind as *housed* in the body; of the senses as the *portals* to the mind, the eyes as *windows* of the soul, the senses the *doors* of perception, and so forth. In these cases it is important to the metaphor that the container is not sealed; but still it is *owned*, entry or exit is restricted or controlled in various ways, and so forth. (The aptness of Leibniz's example of the mill is owed partly to

membership in this family. A mill is like a house which is clearly occupied by its own internal machinery; so we naturally take this machinery as metaphor for the brain.)

Metaphors from this family appear in very many contexts, as when we say that someone who has failed to keep something concealed has *spilled the beans* i.e. let them spill out of his mind/container, and in a way that makes them difficult or impossible to replace. They are, however, particularly common in our conceptualisation of emotion and feeling. Thus we speak of people as *full* of feelings of all kinds, which may *bubble up*, *well up*, or *overflow*, unless they are kept *contained*. We take it that if a person's feelings are *bottled up* then he or she should perhaps seek to *express* them, or *let out* in one way or another, say by *channelling* them into to an activity like art or exercise, *venting* them by talking to an acquaintance, or eventaking them out on the cat, or something of the kind.30 Otherwise the ensuing *pressure of feeling* may be damaging or dangerous.

The family has a number of detailed and systematic variants. Thus, for example, we seem to conceive certain emotions as *fluids* in the mind/body container. We think of anger, for example, as a hot fluid, so that the feelings of someone who is angry are *agitated* and may *seethe* or *simmer*. A person who is *hot under the collar* in this way may be *fuming* as the anger *rises*, or *wells up* in him; and so he may have to *simmer down*, or *cool down*, so as not to *boil over*. If he can't do this, and doesn't somehow manage to *let off steam* he may be at risk of *bursting with anger*, or *exploding with rage*. Here the spectrum of feeling between calmness and uncontrollable anger is represented relatively strictly in terms of the temperature of the emotion-liquid, which may be cool (no anger), agitated or hot (some degree of anger), or boiling (great anger); and the pressure caused by the emotion-heat may ultimately cause the mind/body container to burst, releasing the now vaporised fluid in a sudden, uncontrolled, and perhaps disastrous way.

In some instances of the container metaphor the mind/body container appears more explicitly as the body. Thus we may say that someone is *hot-headed*, or that something *made one's blood boil*. Here the metaphorical emotion-fluid is equated with the blood.31 This also occurs in the case of fear, which we represent in a way opposite to anger (and lust and love)31, that is, by the idea of *cold*. Something which *strikes* fear into one may make one get *cold feet* or make one's *blood run cold*, so that, in the extreme case, *cold fear* or *icy terror* may render one *frozen to the spot* and so unable to move.

Again, where we take the mind/body as a container for feelings, we also take it that significant events may affect the container itself. This happens when we become frozen to the spot, for in this case the coldness of the contents of the mind/body container are represented as affecting, and hence immobilizing, the bodily container. Again, just as another's words may *convey* things from their container to ours, so their words or deeds may *strike* us, and they may *penetrate*, *pierce*, *perforate*, *stab*, *cut*, *sting*, *lacerate*, *lash*, or otherwise attack or injure the mind/body container. Also the mind can be entered, or threatened with entry, in other ways which are connected with bodily entry, as when someone *gets up our nose* or again *bugs us* like some insect intruding on the body. Again, the container can be put at risk from within, as when someone *blows his top, flips his lid,* or *blows his stack*. We characteristically think of such *eruptions* as temporary; but there are more lasting and serious cases of damage, as when someone *cracks up*. Hence also the mind/body container can be damaged, as when a person is *crushed*, *shattered*, *broken*, or *cracked*, and may be unable to recover.

Our taking the mind/body as a container has a further aspect, which is that we liken good things to things which we would like to *put in* to the container, particularly by eating, and bad things to those which should be kept or put *out*. Thus, in general, we regard good things as *sweet*: life is sweet, youth is sweet, peace is sweet, and so, according to our way of speaking, are hope, freedom, victory, revenge, nothings whispered in the ear, people's faces, a moment's relief, dreams, babies (whom one could sometimes just *gobble up*, because they are so *sweet* and delightful), children (particularly little girls, who as we know, are made of *sugar* and *spice*), young animals, melodies both heard and unheard, and an endless variety of other things.

The notion of *taking in* to the mind/body container links with the uses of *seeing* and *grasping* for intellectual matters previously considered. In understanding or grasping things we put them *into* categories, which we also represent as containers (compare the use of Venn diagrams). If a thing is *put in* one container/category, and that is *in* another, then the thing is also *in* the third; so the transitivity of abstract relations of containment in categories, sets, extension of predicates, etc. correspond to that of concrete relations of containment. If one takes in ideas which are *food for thought*, they can *nourish* or *strengthen* the mind/body container. Also, of course, we can take things into the mind/body container in ways other than via the mouth. Just as we can use the eyes to *catch* something, *pick* something *out*, or *hold* something, so we can use them to *take things in* -- even to *drink in* a scene, or *visually devour* something which particularly interests us. Again, we can *breathe in* an atmosphere of fear, or simply *absorb* information, knowledge, the way to perform a task, etc., in ways we are not aware of.

The comparison between things which are good or desirable and things we would like to eat is especially striking in the case of love and lust. Terms of endearment include numerous variations on *honey*, *sugar*, *sweetness*, and the like; she or he may be *the cream in my coffee*, *the sugar in my tea*, my *sweetie-pie*, and so on. One may *hunger and thirst* after righteousness, but also after sexual contact, for which people sometimes say they are *starved*; this extends to all kinds of things, for example to kisses, which may be *sweeter than wine*. There are also variants of these expressions which extend to coarser *appetites*, such as that for *cheesecake* or *beefcake*; and one can want to meet a *dish* or a *hunk*, or take an interest in very many other bodily things compared to food. An extensive comparison between eating and sex seems also to be a part of many languages and cultures.

In addition to representing good or desirable things as things to be put into the body/mind container, we represent bad things as of a kind to be expelled or kept out. This is particularly striking in the intellectual sphere, as it applies to truth and falsity, or to correct or misleading representation more generally. Thus we hold that false or misleading things which someone says may be *trash*, *rubbish*, *garbage*, or even *horse manure* or *bullshit*; and we are liable to say that views we find repugnant *stink*. We may call someone who is fluent at a certain kind of misrepresentation a *piss-artist*. If someone regularly engages in bragging or other relative harmless and self-inflating misrepresentation will may say that they are *full of hot air*. In more serious matters, however, we may say that someone particularly prone to pernicious misrepresentation is *full of shit*. We may find opinions which we characterize in this way *nauseating*; and if so we hope that no one (or at least no one who is not some kind of *sucker*) will be inclined to *swallow* them.

Unfortunately, of course, people may *imbibe* such falsehoods in childhood, or be *fed* them through propaganda, in which case they may be intellectually *infected*, *corrupted* or *poisoned*, as perhaps the young among the Nazis were, without being able to do anything about it. *Contaminating* others with ideas of this kind is decidedly not giving them the proper *intellectual nourishment*. Hence if we find people *airing* such *unsanitary* views, or trying to force such *junk* into our minds, or *down our throats*, we may tell them to *shove it*, thereby indicating that they should put these things back into the inner space of their own body/mind containers, and by a route which reflects their nature.

VI

This brief survey strongly suggests that we should regard metaphor as an important and basic mode of thinking. There are also more general reasons for holding such a view. First, the notion of the comparison of one thing, or set of things, to another, seems a particulary basic operation of thought. When we bring something under a concept, for example, it seems we thereby tacitly compare that thing to the others which we have so categorized, and that this comparison is part of the substance of classificatory thought. Conceptual metaphor seems another way of effecting such comparisons. Also if, as connectionists hold, we think via the activation of neural prototypes of the objects of our thoughts, then we should expect metaphoric processes to be particularly important. For such processes would seem to enable the brain to use existing prototypes for thinking about new domains, and thereby to extend the range of its activity. And we might expect that the process extending the scope of established prototypes via cognitive metaphor would be expecially important in an evolutionary context, since evolution often works by making further use of structures and features which are already present.

Metaphor also shows another feature linked with thought, which Lakoff calls the *Invariance Principle*. This is the principle in accord with which we naturally tend to map source and target domains in a way which is consistent and coherent, so that cognitively extraneous features of the source do not license inferences about the target which are false or absurd. As Lakoff puts the idea

Metaphorical mappings preserve the cognitive topology (that is, the image-schema structure) of the source domain in a way consistent with the inherent structure of the target domain.

Since, as we have seen, metaphorical thinking proceeds tacitly and unconsciously, this constraint is not effected by deliberative reflection; rather it can be compared to a rule of grammar, or of unconscious cognition more generally. Lakoff describes a corollary of this as follows:

...inherent target domain structure limits the possibilities for mappings automatically. This general principle explains a large number of previously mysterious limitations on metaphorical mappings. For example it explains why you can give someone a kick, even if that person doesn't have it afterward, and why you can give someone information, even if you don't lose it. This is a consequence of the fact that inherent taget domain structure automatically limits what can be mapped. For example, consider that part

of your inherent knowledge of actions that says tha actions do not continue to exist after they occur. Now consider the ACTIONS ARE TRANSFERS metaphor, in which actions are conceptualized as objects transferred from an agent to a patient, as when one gives someone a kick or punch. We know (as part of target domain knowledge) that an action does not exist after it occurs. In the soruce domain, where there is a giving, the recipient possesses the object given after the giving. But this cannot be mapped into the target domain since the inherent structure of the target domain says that no such object exists after the action is over. The target domain override in the Invariance Principle explains why you can give someone a kick without his having it afterward.32

We have also seen that we compare the mind to a container or inner space in many ways. So let us now consider the role of this metaphor in relation to the problem of consciousness in more detail.

VII

We noted in section II above that the problem seems based on the intuitive conviction that consciousness presents us with aspects or properties of experience which are subjective and non-physical, and hence beyond the scope of physical explanation. Wittgenstein questioned this; and even if we admit that very many share this conviction, it seems relevant to ask whether there is good reason for doing so. We do not, after all, generally assume that intuition is infallible. To take what seems to me to be a comparable case: In standing still we have a natural intuition of immobility, on the basis of which we infer that we and the earth about us are not moving. This intuition was once accepted without question by very many people, and was integral to the pre-Copernican world view. We now acknowledge that this intuition is misleading; and we also know that it was particularly insisted upon during the period in which the view of which it was an important part was being superceded. Why should the same not hold for the intuition of difference, which, as it happens, is also integral to the Cartesian view of the mind now being superceded by physicalism? It seems to us that we are spatially immobile, when really we are not. So why should it not also seem to us that we are introspectively aware of non-physical qualia, when really we are not? What reason have we for excluding this possibility?

Students of consciousness often emphasize that no real account or explanation of the central notion of qualia, or the subjective or phenomenological aspects of experience, is possible. Thus as Block says, 'the best that can be done is the offering of synonyms, examples and one or another type of pointing to the phenomenon'.33 (The notion of pointing in this context is of course metaphorical.) Accordingly, McGinn offers a caution with nearly explicit relevance to the intuition of difference. As he says, 'We have a sense of the problem that outruns our capacity to articulate it clearly. Thus we quickly find ourselves resorting to invitations to look inward, rather than specifying what it is about consciousness that makes it inexplicable in terms of ordinary physical properties.'34 If even those who take the intuition of difference most seriously tend to regard it as hard to explain or justify, this is surely a further reason for critical examination of it. For this is just what we would expect if the intuition were illusory.

Although the lack of justification for the intuition is striking, students of consciousness often seem unwilling to consider that either justification or explanation might be relevant. Here is a discussion of the issue by Chalmers:

It seems to me that we are surer of the existence of conscious experience than we are of anything else in the world...I find myself absorbed in an orange sensation, *and something is going on*. There is something that needs explaining, even after we have explained the process of discrimination and action: there is the *experience*.
True, I cannot prove that there is a further problem, precisely because I cannot prove that consciousness exists. We know about consciousness more directly than we know about anything else, so "proof" is inappropriate. The best I can do is provide arguments wherever possible...There is no denying that this involves an appeal to intuition at some point; but all arguments involve intuition somewhere...To me, it seems obvious that there is something further which needs explaining here...

Chalmers here writes as if what is controversial -- what he cannot prove, but has to take on intuition as the foundation of his arguments -- is that consciousness, or conscious experience, exists. ('*something is going on*', 'there is the *experience*', etc.) But since, as he says, no one doubts this (and certainly no one who seeks to indentify consciousness with neural activity), these affirmations are beside the point. What might be questioned, of course, is the particular construction, or interpretation, which he and others place consciousness, namely as a veridical manifestation of properties which transcend the physical. For this construction, however, Chalmers offers no reason, beyond that it 'seems obvious' to him, and that our knowledge of consciousness is particularly 'direct' (another metaphor, which might be supposed to reassure us that we need not worry about misconstruing something here). Chalmers' claims thus provide no justification of the intuition of difference; rather they continue a long-established tradition of not examining the intuition, but simply taking it at face value.35

To lack any explanation of the intuition of difference is to lack any solid grounds for holding that it is correct as opposed to illusory. So it seems worth noting that the intuition can be related, in something like an explanatory way, to the metaphor of the mind as a container. To begin to see this it will be useful to consider an example, say the feeling of toothache. In this we feel the pain -- and so, as we can say, we detect the phenomenological quality, subjective property, or qualia of the pain -- as in a space. We feel the pain as *inside* the aching tooth. But of course if we look at the aching place we cannot see the pain in it; if we touch the aching place with a finger, we do not feel the pain with the finger; and the same would be true -- the pain would likewise remain undetectable -- if the tooth were x-rayed, scanned, or whatever. We thus feel the pain as in a space at least partly bounded by the aching tooth, and hence the body; and we also represent this space in a distinct way from that in which we see, touch, or otherwise detect physical things.

Here, it seems, we find something like a genuine application for the notion of two spaces, as used for example by Metzinger in describing consciousness in section IV above. This in turn lends a degree of non-metphorical substance to a number of other notions which we regularly find in the description of consciousness. The sufferer, as we say, feels the pain *from the inside*, that is, from within the space in which the pain is located, and as opposed to others, who are outside that space; and so also the sufferer

can be said to feel the pain in a uniquely *direct* way, and in this to be aware of its *inner* (or subjective or phenomenological) aspect. And finally, since this inner aspect seems manifested in a space which is *not* the physical space in which we see and touch things, it already seems distinct from anything physical. So as we can see, something like the intuition of difference already seems implicit in the representation of the inner space in which pain is manifest.36

To continue with this representation, we may also take it that no one else -- no one other than the person who has the pain -- can possibly be aware of pain in *this* inner space, and so of *this* pain. So it also seems as if the space in which we feel the pain is inaccessible to any means of detection besides the introspective awareness of the particular person who has the pain. Hence it can seem, as Metzinger says, that each of us is aware of such a private space, or 'mental world', in feeling pain or apprehending other phenomenological qualities. And it is the properties or qualities which we apprehend in this way -- that is, in a quality space which is not the space in which we see and touch physical objects -- which seem so intrinsically disparate from the physical as to be inexplicable in physical terms.

It thus seems that this kind of representation of the mind might well might bear some explanatory relation to the intuition of difference. An explanation in these terms, however, would not be a justification. For from a physicalistic perspective the idea that we feel pain in a space distinct from the public one in which we see and touch things involves (as Strawson says in another context) a non-sequitur of numbing grossness. We have reason to hold that pain results from the working of a neural system which serves to make us aware of (potentially damaging) physical events which happen in our bodies, that is, in a certain part of physical space. On this view we might well expect, as seems to be the case, that this particular neural system would not represent visual and tangible things as such, nor show the events or portion of physical space with which it deals as visible or tangible; for, among other things, the events with which this system is concerned occur inside our skins, and hence in a volume of space we do not normally search by sight or touch. So the intuition that in feeling pain we have access to a non-physical space -- as opposed to a physical space which our brains economically omit to represent as having visible and tangible features -- may be just an illusion produced by the working of this particular neural system. (Dominated, as it were, by our visual image of physical space as the locus of visible and tangible objects, we misconstrue our non-visual representation of a bodily space which is out of sight as the presentation of a non-physical space to which we have a peculiarly intimate form of quasi-visual access.)

It seems a real possibility that we should misconstrue a representation of the physical space inside our bodies in this way. But if we acknowledge this, then we should also acknowledge the possibility that our conception of the non-physical internal aspect of pain, which appears as something within this supposedly non-physical space, might likewise be illusory. So on an account of this kind the intuition of difference for pain, like the notion of a non-physical space in which pain is felt, might be a cognitive illusion, produced by the natural working of the brain.

This is a kind of possibility which those who base their conception of consciousness on 'the offering of synonyms, examples and one or another type of pointing to the phenomenon' thereby ignore. Still, since it flows from an approach which offers at least some hope of a deeper account of the intuition of difference, we should consider it further. So let us begin by reviewing Wittgenstein's critique of the metaphor of the

mind as a non-physical space or container in more detail.

VIII

Wittgenstein explains his approach to the 'picture' of the non-physical inner object in the non-physical inner space by saying that 'the best I can propose is to yield to the temptation to use this picture, but then investigate how the *application* of this picture goes.' (?374) This investigation contains two well-known discussions, concerning, respectively, the communicability and the objectivity of descriptions of experience. We can represent these briefly as follows.

First, this picture apparently entails that the subjective aspects of experience are 'private', in the sense of knowable only by the subject of the experience. For if we could know of the subjective aspects of experience only through introspective awareness of our own experiences, and no one could have such awareness of the experiences of others, then no one could know of the subjective aspects of the experiences of others. The subjective aspects of experience would then be, as Crick and Koch aver in the article by Chalmers quoted above, 'impossible to convey to other people.'37 As this quotation illustrates, such a conclusion is frequently drawn but rarely taken seriously. For if it were so, descriptions of the subjective aspects of experience could convey no information about those aspects, and therefore would have no use in communication. (If the aspects really were impossible to convey, then attempts to convey them would convey nothing.) So, reversing the argument, if our descriptions of the subjective aspects of experience do succeed in conveying information, they are not really about something private. Since our descriptions of the subjective aspects of experience serve to communicate information about experience, we must take them to be about aspects of the public world, such as the bodily events which in fact cause and shape our descriptions of experience. And if despite this the descriptions *seem* to be about something 'private' -- something like Wittgenstein's beetle in the box -- then this must be a misconception of their actual working.

Secondly, in our conception of experience we relate the subjective and the objective in a particular way. In speaking of the subjective aspects of experience -- the way pain feels to us, the way colour looks to us, and so forth -- we are speaking of *how things seem* to the persons who have these experiences. So these aspects are literally subjective, in the sense that they are constituted by how things seem or appear to the subject of the experience. (Insofar as it really *seems to me* that my experience is one of pain, say, then my experience *has* the subjective aspect of pain.) This is the notion of subjectivity to which Nagel, for example, appeals, when he writes:

...Very little work has been done on the basic question (from which mention of the brain can be entirely omitted) whether any sense can be made of experiences' having an objective character at all. Does it make sense, in other words, to ask what my experiences are *really* like, as opposed to how they appear to me? We cannot genuinely understand the hypothesis that their nature is captured in a physical description unless we understand the more fundamental idea that they have an objective nature (or that objective

processes can have a subjective nature.)

Subjectivity nonetheless presupposes a kind of objectivity, which requires to be explained. The situation is not, as Nagel might seem to imply, that in considering experience we can dispense with the notion of objectivity altogether (and with it reference to the brain), and consider merely how things seem to us. Rather insofar as we hold that experience has subjective features -- that how our experiences seem to us is how they actually are -- then we are already assuming that these subjective features are also objective, in the sense that they are features which experience not only seems to us to have, but really does have. Hence we hold that our judgments about the subjective aspects of experience are also objective, in the sense that when we categorize an aspect of experience by applying a word or concept to it, we not only seem to ourselves to categorize the aspect correctly, but we actually do categorize it correctly. This objectivity, or more-than-seeming correctness, is essential to our ability to regard experience as constituted by how things seem to the experiencer. For if we could not actually use a concept S correctly, then the fact that it seemed to us that some experience fell under S would establish nothing about the nature of that experience.38 So our notion of the subjective features of experience requires that our categorization of experience be not only subjective, but also objectively valid.

It follows that if we are to be justified in holding that experience has subjective aspects, we must also be justified in holding that we can categorize experience correctly. If experience were private, however, it seems there could be no such justification. To say that our uses of words or concepts in describing experience are objective is to say that these uses can be assessed as correct or incorrect, as by reference to rules or standards, so that a distinction can be made between the case in which a use is actually correct, and that in which it is not, but merely seems so to the user. But nothing could possibly be brought to bear upon the assessment of an application of a word or concept to an individual's private experience besides how that experience seems to the individual whose experience it is. So an individual attempting to apply words to experiences which were private could have no entitlement to claim that these words were applied correctly, as opposed to applied in a way which seemed correct to him, but was not actually so. If experience were private we could have no possible grounds for holding that our categorizations of it were more than seemingly correct. Hence we could have no justification for holding that the requirement of objective categorization was actually met, and no justification for taking experience to have determinable subjective aspects.39 Since the conception of experience as private renders unjustifiable the assumption of objective categorization that it also requires, this conception must be rejected as unacceptable.

IX

It thus appears that if we are to regard our conception of the subjective aspects of experience as justified, we must be able to explain how the requirement of the objective categorization of experience is met. The mere postulaltion of a realm of subjective items -- items whose real nature is whatever it seems to us to be -- clearly offers no such explanation. If we adopt a familiar (Davidsonian) physicalistic perspective, however, it seems that we can construct the kind of account which is required.

In this perspective we ascribe experience to one another together with a great range of other mental states and events, as part of our natural interpretive understanding of human behaviour. Such ascription enables us to perceive (interpret) verbal behaviour as speech, and non-verbal behaviour as action, stemming from desires, beliefs, emotions, and the other motives which serve to render speech and other action intelligible. In everyday life we thus find meaning and motive in one another's utterances and actions spontaneously and continually; and in this we manage naturally to interpret one another with remarkable precision and accuracy. The capacity for such understanding evidently develops together with that for using langugage from early childhood; and we can regard both as parts of a natural system for mutual understanding, produced by evolution, and programmed, in Chomsky's phrase, to 'grow in the brain'. This is also a form of natural and tacit causal explanation, for in making sense of behaviour by reference to perceptions, motives, etc., we are also specifying its physical causes.40

In such interpretive understanding we use language in a particular way. We speak of the desire, belief, hope, fear, or whatever *that P*, where 'P' can be replaced by any suitable indicative sentence. This use of language turns our limited vocabulary of words for motives -- such as 'desires', 'believes', 'feels', etc. -- into a potentially infinity of distinct sentence-embedding descriptions of motive; and such sententential description implements our conception of the mind as having intentionality, that is, as having contents which both represent, and engage causally with, the world. A motive described by a sentence 'P' is thereby specified as related, both normatively and causally, to the situation in which 'P' is true. Thus a desire described by 'P' is thereby registered as one which will be satisfied just if P, and we understand this as the situation which the desire will bring about (cause) if it directs action as it should; likewise a belief described by 'P' is one which is true just if P, and this is the situation which the belief will track (causally covary with) if it operates as it should; and so on.41

Our *that P* mode of description thus yields a form of causal explanation in which we tacitly trace the role of motivational causes via our apprehension of the truth-conditions of the sentences we use to describe them. So far as we can cogently understand motive and action in this way, therefore, we find that the causal operation of our motives is structured in accord with the norms of truth for our sentences. This in turn means that when we understand others in this interpretive way, each of us systematiclly re-finds the norms of truth of the sentences of his or her *own* idiolect in the patterns of motive and verbal and non-verbal action instantiated by the *other*. Insofar as we can understand both our own case and that of the other in this way, we thereby describe thought and action as unfolding in accord with the same norms in both. So this form of understanding not only makes us intelligible to one another, but also encourages us in the normative harmony which sustains our co-operative form of life.

For us to understand one another in this way, however, it seems that certain overall conditions must be satisfied, which Wittgenstein also seeks to elucidate (?206-7) In interpreting others we particularly rely upon their use of language. Speech seems a kind of action which we can interpret with unusual clarity and certainty; and it is through understanding speech that we attain precise and extensive understanding of the experiences and motives of others. Also, however, speech is a kind of behaviour which we could not understand in isolation from the the rest of the behavioural order of which it is a part. If we could not see people's productions of sounds or marks as part of a larger pattern of action and relation to the environment, we could not interpret these sounds or marks, or regard them as language at all. (?207; we

can get a sense of this point if we imagine trying to interpret radio broadcasts of foreign speech, without, however, being able to know what the programmes are about.) By contrast, we can understand a lot of non-linguistic behaviour without relying on language, at least up to a point. We can generally see the purposive patterns in people's behaviour in terms of their performance of commonplace intentional actions, and their being engaged in various everyday projects -- 'the usual human activities' which constitute 'the common behaviour of mankind' and which therefore provide 'the system of reference by which we interpret an unknown language' (?206) But unless we can link such actions with language, we cannot, in many cases, know the precise contents of people's thoughts or experiences; and in the absence of language it would be doubtful how far we could ascribe precisely conceptualized thoughts or experiences to people at all (Cf. ?25, ?32; and also ?342.42

We thus have a general claim about interpretation and understanding. Words with no relation to deeds are unintelligible, and deeds with no relation to words are inarticulate. It follows that the kind of understanding of people which we actually attain, in which we take deeds to spring from motives with determinate and precisely conceptualized content, requires that we integrate our understanding of verbal and non-verbal behaviour, and hence that we correlate and co-ordinate the two. It is some such integration which enables us to tie the complex structure of utterance to particular points in the framework of action and context, and thereby to interpret language; and this in turn enables us to interpret the rest of behaviour as informed by experience and thought which, like that expressed in language, has fully articulate content. I think that the particular mode of integration which we use involves what we can regard as a process of interpretive triangultion. In this process we systematically relate the motives which we take to be expressed in speech -- including desires, beliefs, and experiences -- to those upon which we take speakers to act. We thus triangulate from speech and action to focus upon their common causes, that is, motives (including experiences) which can be specified by relation to both verbal and non-verbal behaviour. In this, therefore, we constantly and tacitly cross-check the motives we assign via speech against those we assign via non-verbal action; and this constitutes a powerful empirical method of interpretation.

The point can be illustrated with a simple kind of example. Suppose that I competently frame hypotheses as to the experiences and motives upon which you are presently acting, and also hypotheses about what the sounds in your idiolect mean. Suppose also that you now make sounds which, according to my understanding of your idiolect, constitute self-ascriptions, or other authoritative expressions, of the experiences or motives upon which I am taking you to act, and your further behaviour bears this out. Then questions of sincerety aside, this tends to show that my hypotheses about both the meanings of your utterances and the motives for your present behaviour are correct, and also that you have first-person authority about these things. So if I could do this for most or all of your non-verbal actions, then I could attain a high degree of confidence about the hypotheses which constitute my understanding of the contents of your motives and experience, and also about your possession of first-person authority. In this, moreover, everything would be confirmed empirically, so that I was taking nothing on trust. My confidence in my interpretations would be owed to their success in explaining and predicting what you did and said, and my confidence in your first-person authority would be based upon its coinciding with my own idependent understanding of the utterances and actions which expressed it. The same, of course, would hold for your understanding of my utterances and actions. And since whatever confirmation each of us had for the interpretation of the other's non-verbal action could be made (via the further

interpretation of self-ascription) to count in favour of each's understanding of the other's idiolect, we could regard our possession of mutual linguistic understanding as confirmed to a particularly high degree. (The principles illustrated here apply to more complex cases.)43

Triangulation of this kind presupposes an intepreter with a capacity to think in a effective hypothetical way about the motives and experiences which explain both verbal and non-verbal actions, and an interpretee who can provide both non-linguistic and linguistic behaviour, where the latter accurately expresses, and so serves to specify, the motives which explain the former. Given these materials, it seems, an interpreter could come to understand the contents of an interpretee's motives and experience with a high degree of accuracy. In the process, moreover, the interpreter could continually check both her own ability to interpret and the first-person authority of the interpretee, and so continually explore and confirm the presuppositions of successful interpretation of this kind. Of course in practice we cannot always interpret accurately, and our first-person authority may fail. But an interpreter can still correct bad interpretations in light of the evidence which the interpretee provides, and also check and, if relevant, try to correct or make allowance for failures in the interpretee's authority. All these processes admit continual repetition and refinement. So the fact that each of us is *both* a potentially accurate interpreter *and* a potentially authoritative interpretee would appear to allow us to calibrate our interpretations of verbal and non-verbal behaviour continuously and cumulatively, and so as to give both something like the degree of precision and accuracy which we observe them to enjoy.

On this line of thought it is no coincidence that we should both possess first-person authority and also be able to interpret one another as accurately as we do, for these phenomena are interconnected. Taken this way first-person authority does not seem solely or primarily a form of self-understanding. Rather it appears as a complement to the ability to interpret: the ability to manifest the kind of correlation between utterance and action which makes precise and fully grounded interpretation possible, and thereby to make oneself understood.44 This dovetailing of abilities, moreover, seems such as to have been shaped by evolution. Other things being equal, we should expect that an increase in the ability to understand and anticipate the behaviour of others should confer reproductive advantage upon members of a species who possess it; and the same should hold for an increase in the ability to influence the way in which one is so interpreted by others, that is, the ability to make onself understood.45 So we might expect that once evolution gets it hands on animal behaviour which expresses motive, it will cull and save in favour of both these abilities. There is reason to hold that such a process has been accelerated among the social primates, and particularly in our own species, and that it has resulted in the interpretive and expressive abilities that underlie our conceptions of language and mind.46

An account along these lines would apparently make it intelligible both how our particular notion of subjective experience should have come about, and how its intellectual requirements are actually met. The precision of interpretive triangulation from verbal and non-verbal effects to their motivational causes depends upon the extent to which the effects themselves can be correlated, that is, the extent to which interpretees can accurately and fully describe or express the inner causes of the interpretable behaviour in which they engage. So it seems both a conceptual and an evolutionary presupposition of our understanding one another in this way that we should have a high degree of first-person authority, and hence a capacity accurately to describe, *inter alia*, how things seem to us, or how things are with us from

our own points of view. It is also in the nature of such understanding that we should constantly check and test one another's first-person authority, both for internal explanatory consistency, and also against our independent interpretive explanation of non-verbal action and expressive behaviour generally. Accordingly, we have a conception of experience in which we take ourselves both to be authoritative sources of information about what things are like for us, and also capable of description of such matters which is objectively correct, and which others can check in a number of ways.47

This sort of physicalistic sketch seems coherent and potentially explanatory. We can take Wittgenstein's discussion of the metaphor of the mind as a container as indicating how understanding of this kind tends to be supplanted by a different (Cartesian) account, which despite its incoherence is somehow so natural to us that we can scarcely free ourselves from it.

On the physicalistic account pain is a neural event, already partly located by interpretation as the common cause of the self-ascriptions and other verbal and non-verbal behaviour which, in any particular case, we take to manifest or express pain. A pain is thus a public physical event which engages with our norms and practices in using the word 'pain', and also an event which occurs inside a container which we do not penetrate in order to determine the reference of that word. As have seen in sections IV - VI above, we also natually think in terms of a metaphor of the mind as a container, and so represent pain to ourselves in this way as well. The metaphor has an impressive degree of fit, because pain occurs out of sight, within the container of the body, and also because we have first-person authority about pain, as we would if it were concealed in a container into which we alone could see. But since the metaphor does not represent the inner space of the body as *such*, once we apply it we automatically *cease* to represent this event as the bodily event which it is, and instead construe it as something non-neural and private.

This is apparently a further characterization of the mistake described as a non-sequitur at the end of the previous section. In using this 'picture' we represent a real inner physical (neural) space by imagining an unreal quasi-physical space which, however, we take as a further reality. Wittgenstein's idea seems to be that having imposed this representation, and so in effect having ceased to focus upon the actual physical event which engages with our interpretive and descriptive practices, we nonetheless *think* we are in direct contact with a 'sensation itself' for which we might frame an ostensive definition, even if we '...only had the sensation.' (?256) Conceived in terms of this metaphor the sensation also appears from outside as 'The thing in the box [which] has no place in the language game at all; not even as a *something*: for the box might even be empty...'. Yet of course we do not actually suppose that the head of a person in pain might actually be empty, or that the role of events in the brain in mediating the causes and effects of pain might somehow be cancelled out. So here, on Wittgenstein's account, we see the creation of a philosophical paradox. The public and bodily event which both self- and other-ascription serve to locate as the reference of 'pain' is represented as beyond the reach of language, and hence beyond the reach of thought, simply by being conceived in terms of a metaphor for its bodily container.

Putting the matter this way seems to come close to the view expressed in some of Wittgenstein's most compressed remarks.

296. "Yes, but there is *something* there all the same accompanying my cry of pain. And it is on acount of

that that I utter it. And this something is what is important -- and frightful." -- Only whom are we informing of this? And on what occasion?
297. Of course, if water boils in a pot, steam comes out of the pot and also pictured steam comes out of the pictured pot. But what if one insisted on saying that there must also be something boiling in the picture of the pot?

It is not to be doubted that there are real events on account of which persons express pain, nor that such events are important and frightful; so there can be no point in insisting that there is *something* there, etc., as if this were the question at issue. Also there is no problem in talking about these events, for like other of the basic objects of speech they tend to be causes of the utterances which describe them, although in this case, as it happens, causes which are inside our bodies. But once we think of these events as ones which we perceive in an enclosed space which is *not* the body, they come to seem indescribable and incommunicable. Things happen in our bodies which cause the verbal and other behaviour through which we express pain, as things happen in a boiling pot which cause the expression of steam. But in the case of sensation, remarkably, we form a picture of these bodily events as internal to a space, and then insist that they are occurring also in this pictured space. We insist that internal events are occurring not only in the pot, but also in the picture of the pot.48

X

Now it seems that few would want to defend a literal version of the conception of the mind as a space or container which Wittgenstein here criticizes. The temptation to think in these terms may, as Wittgenstein says, be constant and overwhelming; but once the metaphor is made explicit, we cannot really credit it for a moment. Where are such mental spaces or containers -- such 'mental worlds' or 'subjective universes' -- supposed to be? Not inside our heads or bodies, for these are physical spaces containing physical things; and not outside them either. Surely there simply is no such space as, in using this metaphor, we take the mind to be. We may represent things as if there were such spaces, but this should not be taken to mean that there really are any.

But then why should we should we have recourse to this metaphor at all, and why should we be so in its thrall as to be tempted to insist that experience has the kind of privacy discussed above? A straightforward answer seems implict in the discussion so far. We naturally distinguish a particular class of physical events as mental, and naturally describe and conceive these events in special, psychological terms. Metaphor is also a way of conceiving things, and the metaphor of the mind as a space or container is a further natural way of conceiving these particular events. In particular, and as a first approximation, we can see this mode of conception as representing something inner by something inner: the representation of the mind as an *inner space* is a rudimentary way of representing the working of the nervous system *inside the body*. On this account such a mode of representation is so pervasive and tenacious in our thinking about the mind precisely because it is one of our natural, preconscious and prescientific ways of representing the events we regard as mental.

In light of the discussion above it seems that we might expect that we should possess a representation of this kind. We can put forward an account of its origin -- which is here very rough and speculative, but which might nonetheless prove capable of refinement into a respectable set of empirical hypotheses -- along the following lines. We suppose that our nervous systems have evolved to anticipate, among other things, both our own behaviour and that of others; and that in the course of doing this they have come to form progressively more explicit representations of the causes of this behaviour. (There would seem to be obvious benefits in this, since representing the causes of phenomena is a means of predicting them. We appear to be the only creatures who make use of a 'theory of mind', that is, an explicit representation of internal causes of behaviour, as opposed, say, to rules which map behaviour to a variety of environmental causes and correlates. So it may be that we are the only creatures in whom this kind of representation has yet evolved.) We can see that the formation of such a representation must have been a formidable task. The neural causes of behaviour are not just out of sight; they are also *not* like the wheels of a mill, in that their working cannot be rendered transparent to ordinary modes of perception in any case. So, as we can say, nervous systems have had to develop representations of nervous systems *blindly* -- that is, without making essential use of perceptual images which showed neurons or neural networks as disposed in space and having features perceptibly related to the operations they perform. And we may speculate that this evolutionary problem has fostered two developments.

The first is the commonsense psychology sketched above, in which we make systematic use of embedded sentences to represent the neural causes of behaviour as motives like desire and belief. Mapping the causes of behaviour to sentences in this way enables us to represent these causes as dispositions related to the types of enviromental situations which the sentences specify. (Thus desires or goals are, among other things, dispostions to bring about situations of a specified type; beliefs are dispositions to alter desires so as to accord with a specified type of situation; experiences are dispositions to alter beliefs to accord with situations of they type which have given rise to them; and so on.) Since we specify the causal roles of these motives by the sentences which describe them, we grasp these roles via our understanding of the sentences themselves. Reusing our worldly sentences as relational descriptions of motive enables us to represent the causal roles of neural structures semantically, and hence with a minimum of reliance upon a perceptual and mechanical image of those structures, as the conditions of evolution would seem to have required.

Secondly, as it seems, we have also constructed an image of the working of the nervous system which coincides with the first-person authority which the above development requires. In this image we represent events in the nervous system by analogy with preception. In perception we see, hear, or touch things in space, and thereby become aware of their perceptible physical features; and in recognizing the perceptible features of things -- as in looking at the wheels in a mill -- we also apprehend something of their role as causes. So faced with the problem of forming a conception of the events in the nervous system which realize experience, we represent these events as another kind of perception, involving another kind of perceptible features, in another kind of space; and then for these events too we can hold that recognizing their perceptible features enables us to apprehend their role as causes. In the simplest cases, for example, we represent visual experience as a kind of inner seeing, auditory experience as a kind of inner hearing, pain as the inner (and vision-like) detection of the quality of painfulness, and so on. We form a primitive conception of otherwise inscrutible events inside our bodies, by representing these

events as perceptible and therefore intelligible, but occurring in another space.

In this, it seems, we make use of a double metaphor. We show the interior of the body as one or another kind of space or container (visual space, auditory space, the inner space in which we detect pain, or again kinaesthetic sensations, etc.), and neural activity as the perception of one or another kind of event in this space or container. This yields a core conception of experience which is relatively simple and general: experience is the perception (or quasi-perception or introspection) of properties displayed in an inner space. Once we start thinking explicitly about the brain, this becomes an account of the role of the brain in producting experience as well. The neural activity which realizes experience is the creation of the subjective properties and spaces of which, in having experience, we become aware; so it is also the location or 'projection' of the relevant inner properties in the relevant inner property-spaces, and hence the use of these inner properties to present things and properties which we regard as outer. In our own case, on this account, the brain creates inner property-spaces which fit harmoniously with their worldly causes; in other cases, such as that of a brain in a vat, the inner spaces could be created in the absence of proper objective correlatives. But in every case, insofar as we employ these metaphors, this is what experience, and hence the creation or fabrication of experience by the activity of the brain, seems to us to be.

Although these metaphors yield a uniform concept of experience, we apply them in a mulitplicity of apparently inconsistent ways. Thus for example we may represent visual space as a two-dimensional inner display of phenomenal shape and colour, located, perhaps, somewhere behind the eyes. This representation -- as seen for example in the internal coloured patches of sense-datum theories, or more sophistocated modern variants -- seems a relatively straightforward mapping of a two-dimensional visible (physical) picture, and hence a physical space containing such a picture, into the physical space which acutally houses the neural events which realize visual experience. This sort of representation suggests the metaphor employed by Chalmers above, that our acquaintance with consciousness is particularly 'direct': for on this understanding the most immediate object of our acquaintance is the two-dimensional array, which we can perceive with such immediacy and accuracy precisely because it is confronts us *within* the imagined inner space.

Alternatively, however, we may think of visual space as somehow extending out into physical space, as when we take colour as both a feature internal to our minds (or of our 'percepts') and also as a mode of presentation of the surface of an external object. Or yet again we may 'drink a colour in', first thinking of the colour as external to us, then as somehow entering our heads or minds; and so on. This elasticity in our conception of inner space seems a natural consequence of its metaphorical nature. When we think of visual experience as a metaphorical seeing of inner objects we relate source and target domains which overlap50, but which nonetheless remain distinct. The source events involve spatio-temporal and causal relations between our eyes and objects in the environment, whereas the targets are in our heads. So sometimes we stick closely to the intrinsic nature of the targets, and think of the relevant space as somehow in the head; and sometimes we are drawn more closely to the source, and hence think of the targets as encompassing the space and objects of the environment as well. In this way are able to shrink the metaphor towards the target or expand it towards the source, while the underlying mapping remains the same.

XI

The overall empirical hypothesis suggested here is thus that evolution has provided us with two representations of the neural causes of behaviour, which we may call the sentential and perceptual images respectively. In employing the first we ascribe mental states and events via words for motives, commonly together with embedded sentences; whereas in employing the second we think of mental states and events as objects of a kind of internal perception, which is modelled on the real physical thing.51 Both these images represent events or states inside the body by mapping them to events or states outside; and both might be seen as extending the use of neural prototypes (of interpreted sentences, or instances of perception) which already have an independent function. Moreover the ways in which these images represent causal role seem complementary. We use the sentential image in the great range of cases (the propositional or sentential attitudes) in which our internal states can be descritpively related to distal objects and situations, and so can be specified sententially by reference to the environment. In such cases the alternative, perceptual image is often recessive or absent; but it comes to the fore in cases in which something more like demonstrative specification of particular aspects of experience is required.

The claim is thus that we think in terms of a visual inner space, just as we think in terms of a journey of life. In both cases the phenomena are so interrelated as to support the mapping we employ in thinking about them. But since in both cases the thinking is a form of metaphor, there is in reality as little (or as much) such as space as there is such a journey. In the case of the mind, however, the image of inner space seems particularly basic to our thinking; and in consequence we can use the insight that this image is a form of metaphor to explain features of our conception of mind. In particular, we can now explain the intuition which leads us to distinguish the mind as a special realm in the first place, that is, our intuition that there is a profound difference between phenomenal and physical properties.

Suppose it were true that in order to represent neural events inside our bodies -- events which we could neither see nor understand by means of sight -- we had come to portray these events as hidden in a container, and perceptible by another kind of sense (introspection). Then in accord with the invariance principle set out in section IV above, we would be required to represent this alternative container as if it were *not* the physical body, and the properties of these events as if they were *not* perceptible physical properties. For otherwise the metaphor would conflict with the anatomical nature of the target: the metaphor would imply, for example, that we could open the body or head of another and perceive the internal objects or properties, and this is obviously not the case. Thus we can see that the idea that one might look into another's mind/body container, or open it to perceive what is within, is comparable to the notion that someone given a kick ought somehow to possess that kick afterwards. These are both possible (grammatical) consequence of the use of a metaphor, which we do *not* draw, because they are overridden by the intrinsic nature of the target domain.

In metaphorical mapping generally, as Lakoff says, 'inherent target domain structure automatically limits what can be mapped'. In this case the inherent structure of the targets requires us to represent them as both

(i) hidden in a container and (ii) not admitting external sensory understanding in any case. According to the present hypothesis we satisfy these requirements by holding (intuiting) that both the container in which the targets are hidden, and the perceptual properties which we take them to display, are *not* ordinary physical ones. So this, we hypothesize, is the origin of the intuition of difference. This intuition is a result, or expression, of the operation of the invariance principle in our natural use of metaphor in thinking of the mind/brain as an inner space or container.

This can be put particularly clearly in the case in which the visual field is conceived by analogy with a real physical object or picture. Here the target is neural activity in the visual system, and the source is the activity of perceiving something physical. When we conceive the target by mapping the source onto it, we think of visual activity as involving the perceiving of something in the inner space of the mind or head. This risks contradiction with our knowledge that the head is not hollow, that examination of the brain shows nothing like little coloured models, images, or pictures within, and so forth. So in accord with the invariance principle we think of this perceiving as involving a *non*-physical object, in a *non*-physical locus, which is still, however, somehow closely related to the eyes. This is the visual version of the process described with reference to pain at the end of section IX above. But here the process yields the picture of the non-physical soul which sees experience, and thereby, although inside us, sees what we see.

In the case of perception our sentential and perceptual images of the mind/brain overlap. We can characterize an inner event of perception by saying that it is an experience or perception *as if P* (as if I were seeing an orange flower); and again we take it that the experience or perception has an introspectible character, which we represent as an inner seeing of the orange flower, or of its image. Thus using the sentential image we represent the inner event involved in the situation of S's seeing an orange flower via the sentence 'S sees an orange flower', which describes that situation; and in the preceptual image we use a visual image of the same situation, that is, an image of S seeing an orange flower. In both cases, therefore, we conceptualize the inner state or event by mapping it to the outer situation which we take it to be about. On the present hypothesis this kind of intentional mapping is one of our most general ways of representing the mind. The perceptual image can also be regarded as an instance of conceptual metaphor, in which both source and target domains are physical. Since taking them as such renders the image incoherent we automatically regard the metaporically transposed space, objects, and properties as non-physical, and hence as having nothing to do with the body, within which, however, we tend to think of them as somehow located.

The process hypothesized here is simple enough to diagram, and indeed, since this metaphorical way of thinking is so common, pictorial representations of it are already quite familiar. We hypothesize that the problem (taken from an evolutionary perspective) is to develop a characterization of the unknown inner events involved, e.g., in visual perception. We may represent this problem as follows:

[no image]

The solution (again *sub specie evolutionis*) consists in a metaphorical mapping of the process of perception onto the inner events which it is to our selective advantage to be able to cognize. This yields an image in which the targets are characterized as follows.

[no image]

Since this image is at variance with the nature of the targets which we use it to think about, we represent the internal objects of perception as lacking the physical characteristics, such as solidily and postition in physical space, most obviously inconsistent with anatomy (the 'minds eye' is often simply deleted). In thus rendering our image more nearly coherent, however, we perforce begin to think in terms of an internal realm of objects of visual consciousness, which while perceptible are also non-physical, and hence private.

We might use such a diagram to illustrate the hypothesized origin of the conception of inner colour which we find in Chalmer. He expresses the problem of consciousness by saying 'I find myself absorbed in an orange sensation, and *something is going on*. There is something that needs explaining.....there is the *experience*.' What needs explaining about the experience, presumably, is the ostensibly non-physical property of (say) *phenomenal orange* -- the experienced presentation of colour as inner, subjective, and private. On the present account this is to be explained via the operation of the invariance principle on a mapping of the kind shown in the diagram. According to this explanation there really is no inner colour, nor any inner space nor phenomenal property displayed in it; there is nothing which is both real and inner besides what goes on in the body and brain. In thinking of *the experience itself*, however, we use our natural means of thinking about these inner processes, by mapping them to external perception; and so we *think* of the experience in terms of colour and shape which are somehow inside our minds or heads. Just as in an optical illusion we may experience one thing as in a location acutally occupied by another, so in this cognitive illusion we experience colour and shape as in a location actually occupied by the brain. In this case, as in other illusions, the ghostly image we thus produce is in fact derived from things which are entirely substantial. The process is like a conjuring trick, or joke, played on us by our system of representation, in which we mysteriously combine purely physical ingredients -- the employment of a physical metaphor, in the understanding of a physical process -- in such a way as to produce the image of something at once inner and non-physical. This, I think, is the 'logical sleight of hand' which gives rise to the intuition of difference, the supposed non-physicality of the mental, and the problem of consciousness which we have been considering.

XII

Of course very many objections might be made to this account. A first response, for example, might be straightforward incredulity. We are dealing with something which is very familiar to us, namely thinking of our own experiences. The account which we are offered of this familiar activity is in terms of evolutionary speculations which are, to put it mildly, far from grounded in the fossil record; and these speculations concern the development of mechanisms of thought which are obscure, which are supposed to work without our being aware of them, and which function to render our natural understanding of experience quite illusory. What reason is there for taking such an account seriously, let alone for accepting it?

The answer turns upon a particular conception of the data to be explained. Chalmers and others who discuss the problem of consciousness are certainly right to think that something about experience needs explaining. What seems to me to need explaining, however, is not the realm of non-physical properties to which experience seems to give us access, but how we should have come to think in terms of such a realm in the first place. The idea that we have unmediated access to any sort of reality -- that we do not have to go to the trouble of representing this reality to ourselves, but are somehow just *given* it, as in introspection -- already seems suspect; and that such a reality includes non-physical events and properties somehow inside of us does not bear serious examination. So what seems to me to need explaining is why, if things are *not* this way, we should be so determined to represent them as if they were. The account above, far-fetched as parts of it are, provides the outlines of an account of this; and I know no better. In this case the beginnings of an explanation seem to me better than nothing, and better than trying to make a virute of having no explanation at all (but that of course does not improve the explanation itself). It is clear, however, that someone who thinks differently about the data will think differently about this proposal for understanding them. In particular someone who thinks that we really are given an inner realm of non-physical items or properties will think this account not so much an explanation as a denial of the basic data which require to be explained.

So what are the data? Does experience really present us with properties which are non-physical and private, or is the more basic fact just that it seems to us that this is so? Our previous discussion of the intuition of difference strongly suggests the latter. First, as we saw, those who stress the inner manifestation of something non-physical and private do so on the grounds than that this seems obvious to them, or again that it if we consider our experiences it will seem obvious to us, and so forth. (Compare the statements by Block, McGinn, and Chalmers in section VII above). It may well be reasonable to accept how things seem, and it is certainly to be expected that people will do so; but the data which prompt such acceptance clearly do not thereby go beyond how things seem. Secondly we have argued that we can have reason to regard first-person judgments about experience as correct and authoritative only insofar as we can take these judgments to be about the inner causes of speech and other behaviour which are located by interpretation; and these causes seem to be physical. This implies that first-person judgments about *non*-physical inner events and properties should not be taken to be objectively correct; so the most they can register is how things seem to those who make them. Since nothing has been said in favour of the view that the basic data are *more* than seemings, and there is a strong argument that they can be *no more* than seemings, this is the best conclusion to draw.

It seems to me that to regard the data in this way is in fact only to take a minimal step away from the Cartesian view of the mind. The core idea of the Cartesian tradition is that introspection gives us unmediated (indubitable, clear and distinct) access to a reality which is inner and mental, so that the relation of this reality to the physical world remains to be determined. We should see it as an immediate consequence of physicalism that this claim about introspection is false. If there is good reason to hold that mental events are physical, and introspection does not show this, then the first conclusion to draw is that there is good reason to hold that introspection does not after all give direct access to mental reality, but requires to be understood in some other way. So in fact we distance ourselves from the Cartesian tradition only a little if we hold that what introspection gives us access to is not, strictly speaking, an inner *reality*,

but rather *a way things* seem to us with regard to the inner, the accuracy of which might be considered further.

Given the otherwise mysterious nature of introspection this degree of caution does not seem excessive. But once we accept that the basic data concern how things seem to us, the problem of consciousness is transformed. In the first place, the claim that experiences really do have the non-physical properties which they seem to have clearly provides no further explanation of them. In the case of real perception the claim of veridicality coincides with the provision of an explanation: the fact that I seem to see something red, for example, is (very often) to be explained by the claim that there is something red in front of me which I really do see. Here the *explanans* and *explandum* can be separately established, and are linked by causal processes which can be investigated and understood. But the fact that I seem to introspect a non-physical inner property of redness (or as of seeming to see redness, or whatever) cannot likewise be explained by the claim that there really is such a non-physical property in front of me in inner space which I really am introspecting. In this case nothing can be established besides the seeming, and there is no reason (no kind of familiar causal connection, no further results of investigation) to think that beyond this seeming there is a non-physical reality which renders the seeming veridical in this particular respect. In fact the idea that there must be an inner psychic reality to which these inner psychic seemings correspond seems itself just to be a further product of the metaphorical comparison with real perception which is under discussion.

This means that from an explanatory perspective the claim that our introspective seemings are truly non-physical is an addition to the data, which should require further justification. So to treat the problem in the usual introspectionist fashion -- to resort, as McGinn says, 'to invitations to look inward, rather than specifying *what* it is about consciousness that makes it inexplicable in terms of ordinary physical properties' -- is already to take a step which is without warrant in the explanatory terms in which the problem is cast. (At this point, as Wittgenstein says at ?308, 'The decisive movement in the conjuring trick has been made, and it was the very one we thought quite innocent'.) If, however, we ask instead how the *appearance* of non-physicality in this case might be explained, then we can seek to frame a different kind of account, of which that sketched above is one example. Again, taking introspection as giving access to an *appearance* (or representation) of the inner allows us to understand the mind as involving genuine internal representation of internal states and events, and hence to see why we should take mind as inner in the first place. No doubt alternatives of this kind are less exciting than the prospect of a revolutionary new science of consciousness; but they may prove more to the point.

A second objection, less immediate but philosophically deeper, takes us to the distinction between primary and secondary qualities. This objection is, that the present proposal does not resolve, but merely re-locates, the explanatory problem with which we are concerned. According to our proposal the ostensibly non-physical character of the phenomenal property of, say, *inner red* is to be explained by reference to (a metaphorical mapping, involving the principle of invariance, to) the visual perception of red. But this leaves the fact that we perceive things as red itself in need of an explanation. When we perceive an object *as red* the object is thereby presented to us in a certain way, that is, as having a certain experienced visible property; and even if we take this property to be identifiable with, or realized by, some physical property such as reflectance, we must still explain why this realizing physical property is experienced by us, or presented to us, in the particular way it is. This is traditionally done by saying that

the realizing property causes experiences with a certain phenomenal character, in virtue of which we take these experiences as presenting the objects which cause them *as red*. But on the present account, in which the external perception is used to explain the impression of the inner, no explanation of this kind can be given. So we are still faced with a non-physical property -- that constituting the perceptual mode of presentation of visible redness -- which does not in fact fit into our physical world view. The locus of the problem of the non-physical property has, as it were, been shifted from inner to outer; but the problem itself remains unresolved.

To make this objection clearer we can put a particular version of it as follows. It seems (as I think, and as I take the objector to hold) that when we see something red, we see it as *this* (or like *this*) where *'this'* ostensibly refers to a property which we experience (see) the surface of the object as having. Alternatively, however, we can regard this as a feature of the experience itself, for example as the property which occupies the portion of the visual sensation, or region of the visual field, which corresponds to the presented surface of the coloured object (cf Chalmers' 'orange sensation' above.) As we impose our scientific image of the world, we come to think of the surface of the object as having just the property of reflecting light in certain ways; so we think of this as our experiential mode of presentation of this light-reflecting property of the surface, and hence as really located in the mind. We thus take the *this* revealed in experience as inner, and thereby as a manifestation of a non-physical property, and so as providing one more example of the problem of consciousness. But as the objector notes, acceptance of the account given above entails that when we see something red as *this*, we can no longer explain the ostensibly outer property referred to by *'this'* in terms of such phenomenal presentation or projection. So we must take the fact that we see outer red (or such-and-such a reflectance) as *this*, as a fact which is brute and inexplicable. But this brute fact is comparable to the brute fact that pain feels like *this* (with, of course, a different *this*): it is just the problem of the non-physical properties given to us in consciousness, but now in a different guise. If we banish the inexplicable from the inner, it returns to the outer, where it remains unexplained.

This argument can be applied over the whole of the manifest or commonsense perceptual image of the world. The claim that experience gives us access to an unexplained *this*, which we can consider either as inner or as outer, holds not only for how things look, but also for how they sound, feel, smell, and so forth -- for how they are presented to us in experience generally. In each case we seem presented with an experientially detected property which is ostensibly outer, but whose commonsense causal role in perception can be supplanted by some physical property of the object of experience. In reflecting on the causation of experience we therefore construe this experientially detected property in an alternative way, that is, in terms of the relevant portion of the appropriate sensation, sensory field, or whatever, and so as an inner mode of presentation of the outer physical property which has usurped its role. Thus when we take sounds as realized by compression waves, or odours by molecular dispersion, we feel that the question as to what makes the wave a sound like *this*, or the dispersing substance a smell like *this*, is to be answered by reference to the inner *this* which serves to present the realizing physical thing or property to us in experience.

Nonetheless it is clear that we cannot regard this way of thinking as compelling. For also in every case we find (or now assume that we will find) that just as the external role of the *this* of experience has been

supplanted by some outer physical thing or property, so also the inner explanatory role, that of providing a mode of presentation for the outer, is actually discharged by mechanisms in the brain. We do not need to allude to the phenomenal character of visual experience, for example, to explain why we respond to certain reflectances as manifestations of red; for investigation indicates that these reflectances have a physiological salience to us which dovetails with their cognitive and psychological role. So just as explanatory considerations lead us to banish the *this* of experience from the outer word, it seems they ought also to lead us to banish it from the inner as well, insofar as we persist in regarding it as non-physical. It is true that in our manifest or commonsense image of the world we do not think of colour, sound, or odour as we do as a result of scientific investigation, and also that we think of experience as presenting us with such perceivable properties of objects; but this does not mean that we have any explanatory reason to regard the commonsense image as the result of our clothing the expernal world in non-physical properties projected from within.

We can see that this dialectic reflects the claim made in section II above, that the intuition of difference between phenomenal and physical goes with the distinction between primary and secondary qualities. Also the discussion partly recapitulates a phase of intellectual history, in which commonsense properties have been re-understood in terms of the character of experience itself. Roughly, where we have found that the commonsense image of the world requires to be replaced by a distinct scientific one, we have tended to understand the manifest image as an experiential mode of presentation of the scientific, and so in terms of a form of phenomenal projection.52 In this way, as one might say, we have replaced the problem of the manifest versus the scientific image of the world with the problem of consciousness, which we have been discussing here. The objection we have considered, however unscientific, merely threatens to undo this replacement.

But the way of thinking which has led us to this replacement, historically sanctioned though it is, must surely be regarded as mistaken. For this whole discussion, far from constituting an objection to the account above, seems best understood in terms of it, and particularly in terms of the metaphor of containment upon which the account turns. For as we reflect upon our considerations -- of properties first regarded as outer, and then taken into the mind; or again of the mind as projecting non-physical inner properties outside, so as to present to itself an otherwise noumenal external world -- it must begin to dawn on us that we are not really engaging in explanatory considerations relating to natural science at all.53 What we are doing (or so it seems to me) is working through various versions of the metaphor of the mind as a container, by considering properties as located inside or outside this inner space, or going from one location to another, or whatever. The discussion is really an exercise in the exploration of a metaphor; and although this metaphor may be particularly important for our *conception* of mind and experience54, still the metaphorical space *itself* is something which, on reflection, we have every reason to regard as entirely imaginary. So we are not locating things in real space, but in relation to a space which exists only in our imaginations. This is not, on the face of it, an application of science, but at best a mode of reflection upon our system of representation; and unless we make ourselves aware of the nature of the activity, it is likely to be an indulgence of confusion.

This applies particularly to the dual role which discussions of this kind assign to the properties discerned in perceptual experience. This duality -- in which properties seem capable of being regarded as outer, or

alternatively as inner modes of presentation of the outer -- seems itself best understood as an indication that we are thinking in terms of the metaphor under discussion. It is natural that we should do this, but we should not forget that this metaphor is a way of thinking of things in the physical space inside our bodies, that the physical location of such things is the only real location they have, and that metaphorical location, or location within a system of metaphor, is quite another matter. In many cases we have no tendency to become confused about this. If someone says that a telephone number has just gone out of his mind (or head) it would only be a joke to ask whether it got out through the nose or mouth, where in the surrounding room it is located, etc. In this sort of case everybody understands the description as the metaphor it is. But contrast the use of space in the quotation from Metzinger at the end of section III above. This seems to me best understood as an instance of the same metaphor -- much more explicit and full, but metaphor nonetheless. But probably there are those who take it to be literally true, and who think it relevant to engage in research, say, as to how energy or information might be transferred between our private psychological spaces and the physical space in which they are embedded.

If in assigning this dual role to experienced properties we are thinking of them in terms of this metaphor, then we are thinking of them as we would if the account sketched above were true. (See also the end of section X.) We have seen that to think of a property in this way is already to think of it as detachable from reality and non-physical when considered as inner. So it does not seem that the fact that experience seems to present us with a *this*, inner or outer, which is seemingly non-physical, can be the basis of an objection to this account. Our tendency to think this way seems consistent with the account, and at least partly explained by it. Also, the account explains why it should be *perceptually experienced* properties, rather than others, which we think of in this particular way. The properties in the manifest perceptual image of the world are precisely those were available for the kind of evolutionary provision of metaphor which the account hypothesizes to have taken place. The mapping of these properties to the inner to aid in conceptualizing it resulted in their replication in a seemingly immaterial form in the mind/body container. This meant that they were fated to be regarded as really inner after their causal and explanatory role was displaced by science. Thus although our account compels us to regard a part of our intellectual history as erroneous, it seems also to provide some explication of the error, which touches on both its form and scope.

Replies like this might be made to a number of objections to the account presented above. But it is worth noting that such replies should *not* tend to change an objector's intuitive feelings about these matters. Despite formulating the argument above I remain inclined to feel that when we see something as red we see it as *this*, just as when we feel pain we feel it as *this*, and that the properties indicated by this in both case are non-physical, *sui generis*, and inexplicable. Probably the reader feels this too. So in assessing the account it may be worth remembering that according to it this is just how things should be. Such observations are part of the data, and the account is such as to keep these data intact while attempting to explain how they are consistent with physicalism. Our claim is that we do indeed think of the mental in this way -- that is, in terms of a metaphor of the mind as the container of experience, or the space in which experience is perceived -- and that this is as natural and valuable to us as we might expect evolution to have made it. There is no reason to expect that such a deep matter would (or should) be altered by a grain of explanation.

The present proposal thus fits with our natural intuitions about our inner lives, except in regarding one of these intuitions as only seemingly veridical. It accords with our proposal, for example, that our experiences are real and palpable to us, that we seem presented with them particularly directly and vividly, and that we should naturally take them as constituting a realm distinct from any other, whose significance is unique. The proposal is just that what enables us to conceive the inner in this way is partly a system of metaphor whose operation makes the inner *seem* non-physical, when this is not really so. According to this hypothesis we are natural Cartesians; but our innate inclination towards dualsim is explicable in terms of a physicalistic understanding of the world which it constantly tempts us to repudiate. If this hypothesis is on the right lines, then just as we have not been shaped to feel the motion of the outer world, so we have not been shaped to *feel* the physicality of the inner. These are cases in which nature has not enabled us simply to feel that things are as, on scientific grounds, we take them to be. In such cases we can judge things rightly only if we depart from the spontaneity of intuition, and remain content with the discipline of thought.