Representation of the Inner and the Problems of Mind

Jim Hopkins

Three philosophical problems -- the problem of the external world, the problem of other minds, and the problem of consciousness -- seem rooted in the way we conceive experience. We tend to think of our experiences as having a nature which is radically distinct from that of the world which they present to us. This emerges in a series of oppositions as between experience and the world, which we can set out as follows.

1. Oppositions between the inner and the outer.

Our experiences of the worlds seem to be internal to the mind, whereas the world which they present seems external to it. Thus when I feel pain or see a tree, the experience of pain, or the visual impression of the tree (and the visual field of which this experience is part) are internal to me, and hence presented in introspection, whereas the tree itself, by contrast, is part of the external world. The internal character of these experiences, moreover, seems to be phenomenal, and hence not that of any physical thing. This character also seems subjective, in the sense that what it is, is wholly and fully presented in how it seems in introspection; whereas the characteristics of an external physical thing like a tree are objective, in the sense that there is a potentially rich distinction between how they seem in perception and how they actually are. Finally, the internal experience seems private, in the sense that it can be introspected or apprehended only by one person, the person to whose mind it is internal, whereas a physical thing is public, in the sense that in the sense that it can be perceived and hence known about by more than one person.

These oppositions apparently arise early in our cognitive lives. Children of three, for example, already distinguish between a physical item such as a dog and its corresponding visual image, holding that the latter is 'just in the mind', where only one person can see it. Although this presumably reflects the fact that the events we describe as mental -- including visual experience and imagery -- are realized in the nervous system, it nonetheless remains puzzling. For when we consider such events in introspection, their innerness does not *seem* to be that, or only that, of being physically inside the body. A visual image, for example, may seem to be somewhere behind the eyes; but it also seems to have spatial aspects or regions of its own, which are not those of anything inside the skull.

This applies even to events which have a precise internal bodily location, such as pain. We feel the pain of an aching tooth as *in the tooth*, but we also hold that no examination of the physical space occupied by the tooth will reveal the pain itself which we feel there. Rather the felt quality of the pain seems to be in an internal locus which is introspectible only by the person who actually has the toothache, so that this introspectible space seems private and distinct from the public physical space inside the tooth and body. Indeed, as the case of pain in a phantom limb makes clear, this phenomenal space can apparently be

occupied even at a locus at which nothing real exists. And as is familiar, the introspectible quality apparently manifest in this internal space seems to be the defining or essential feature of a sensations such as pain.

Taking the case of visual experience, we can present these oppositions in the form of a diagram as follows:

Figure 1: The dualistic image of the mind.

2. The conceptual nature of these oppositions.

These oppositions seem conceptual, in the sense that they reflect differences in our ways of thinking about the mental and the physical which are both deep-seated and interconnected. These differences seem organized, as presented here, around the opposition between the internal and the external. What is external and externally perceptible is physical, and what is physical is objective and public in the senses above. Indeed the objectivity and publicity of physical things seem to flow from their existing in an external public space, and hence being observable by more than one person and from more than one point of view. By contrast, what is internal and introspectible is phenomenal, and what is phenomenal is subjective and private. Again, the subjectivity and privacy of phenomenal things seems to flow from their existing in an internal private space, and hence being introspectible by just one person, and from just one point of view.

Likewise it seems part of our way of thinking about these matters that no one thing could satisfy both of any of these pairs of oppositions. It seems that no single thing could possibly be internal to the mind in the way experiences are, and also external to the mind in the way physical objects are. Again, it seems that no one thing could be both introspectible and extenally perceivable, or both phenomenal and physical, or both private and public. Hence these oppositions naturally give rise to dualism, the view that mental phenomena are not physical, and hence that human experience is not, as we otherwise have every scientific reason to believe, realized in the brain and nervous system.

These oppositions yield the problems connected with them in relatatively straightforward ways. First, according to this picture, we know the world via our experience of it. Our experiences therefore constitute our ultimate grounds for judgments about the world. It is solely because experience presents us with a three-dimensional world of physical objects, including stones, trees, and other human beings, that we believe that this is how things are. But as the oppositions demonstrate, experience itself is radically different in character from the world it presents. As Berkeley says, nothing but an idea could be like an idea. It follws that the nature of the world must be distinct from , and hence logically independent of, that of experience. But then, taking our experience in general, there seem two possibilities: either the world is as our experience presents it, or not. Since we have only our experience to go on, and since the nature of our experience is independent of that of the world, we can do nothing to determine which of these possibilities obtains.

Secondly, even if we accept that we can know about the external world, including other human bodies, it seems that we can know nothing of the experiences of others. For according to the picture above these eperiences are internal and private, and hence inaccessible to us and not part of the external world in which others are located. This being so, it seems we can know nothing about such experiences, and indeed can have no reason to assume that they exist. For if all that others do -- all the ways their bodies move, for example -- can be taken as caused by events in their brains which are distinct from experiences, why should we postulate anything further? And even if we identify others' experiences with events in their brains, still we seem to have no access to the phenomenal character of those experiences, and hence might still wonder how far they are comparable to ours. And even if we take this as determined by their physical nature, still this determining by the physical and public of something which is phenomenal and private remains mysterious and arbitrary. If this is a consequence of physicalism it nonetheless requires to be explained.

Finally, even if we accept that we can know about the minds and experiences of others, and also accept that these experiences are realized in the brain, still we are unable to explain how events in the brain can have the properties which we take experience to manifest. How can neural events, which are physical, objective, and public, realize or constitute experiences which are phenomenal, subjective, and private? This problem is clearly not solved by accepting that mental events are in fact events in the brain, for the difficulty is precisely that of understanding how this can be so. Hence it is also clearly not solved by holding that neural events can have aspects or properties which are phenomenal, subjective, or private, since this again is a version of what requires to be explained. Nor finally is it solved by supposing that we have distinctive first-person ways of thinking or modes of presentation of the neural events which realize our experiences, although this certainly seems to be the case. For the question remains, as to how and why these ways of thinking should render those neural events phenomenal, subjective, and private, or make them seem so, despite their physical nature.

Philosophers have recently concentrated mainly on the problem of consciousness, and have tended to take this in terms of the opposition between the physical and the phenomenal. Even in recent discussions, however, echoes of the other oppositions and problems remain. Thus consider 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....

Here the focus is not on the traditional problems of knowledge, but a proposed new science of

consciousness. But if we really were such private worlds as this picture suggests it is not easy to see how we would be justified in acknowledging the public world in which we were so mysteriously embedded, nor again the other private worlds scattered throughout it. So we might naturally wonder whether we should accept a picture of our situation according to which public space is like Swiss cheese, bubbled through with little spaces of a distinct kind, awaiting some Einstein of the psycho-physical to chart their alternative geometry.

Partly for this reason, philosophers tend to divide in their approach to consciousness, as between those who take such oppositions as above to be unquestionably real, and those who hold that they are only apparent. The conviction that such oppositions are real leads naturally to the idea that our present scientific approach is fundamentally defective in failing to encompass them, and hence that the understanding of consciousness will require basic revisions to physics, the concept of space, and so on. The idea that such oppositions are only apparent, by contrast, suggests no such revisions, but requires an account as to how these appearances are generated. This, as we noted above, is also what an explanation of the problematic features of experience in terms of first-person ways of thinking of neural events should aim to provide.

3. Mental concepts as ways of thinking of neural states and events.

How might the problematic features of experience be explained as artefacts of our concepts or ways of thinking? To see this we first need to see how we can regard our mental concepts in general as ways of thinking of neural activity. For this purpose let us put phenomenonly aside, and consider instead the way we think about such mental states as desire and belief. We can call these sentential states, to mark the way we conceive and describe them via embeddings of sentences which would otherwise describe the world.

We can set this out by observing that each of us speaks and understands an idiolect, which we take to be that of a natural language. In understanding the sentences of our idiolects we know the conditions in which they are true. So for example each speaker of an idiolect of English knows

The sentence 'Freud worked in Vienna' is true (in my idiolect) just if Freud worked in Vienna.
The sentence 'Wittgenstein lived in Vienna' is true (in my idiolect) just if Wittgenstein lived in Vienna.
The sentence 'The moon is blue' is true (in my idiolect) just if the moon is blue.
and so on.
Each of us thus knows, or can readily come to know, an indefinitely large number of truths relating the sentences of his or her idiolect to objects and situations in the world. We can schematize this by saying that each knows indefinitely many instances of the form:
T: 'P' is true (in my idiolect) just if P
Where occurrences of 'P' might be replaced by appropriate variants of sentences of the idiolect.

In commonsense psychological thinking we re-employ these sentences in a familiar and significant way. In representing states of mind we use a vocabulary of verbs for motives, such as 'desires', 'believes', 'hopes', 'fears', etc., each of which admits of complementation by a further sentence. So we speak of the desire, belief, hope, fear, etc., *that P*, where 'P' can be replaced by a sentence suitable for specifying the

object, event, or situation towards which the motive is directed. This assignment of sentences to motives makes it possible for us to use a finite stock of psychological concepts to generate a potential infinity of ascriptions of mental states, and enables the truth-conditions of the sentences to do duty as conditions of satisfaction, fulfilment, realization, etc., which we ascribe to the motives themselves.

A motive described by a sentence 'P' is sematically related to the situation in which 'P' is true; and this enables us to specify causal relations holding for motives in a sentential and semantic form. 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 functions as it should, and which, again, it will carry into the contents of desire in practical reasoning. This kind of semantic specification of causal role enables us to use our linguistic understanding of sentences, as schematized in **T**, as a mode of understanding of causal relations, as these hold between motive and motive, or motive and world. Such linguistic or hermeneutic understanding underpins our everyday grasp of the working of motive, and we can employ it even in the absence of an explicit concept of causality.

We can think of this understanding -- artificially but usefully -- as built up in two stages. In the first we learn to relate words and sentences to things and situations in the world, and so come to undersand the primary sentence-to-world mappings of our idiolects, as registered in **T**. Then secondly we learn to relate these sentences to our motives in commonsense psychological ascription, together with the situations now linked to them via the concept of truth. It is this further mapping of motives to sentences which enables us to describe these inner causes of behaviour, and their relations with one anothter and the world, in semantic terms. (Individuals who can speak but lack a 'theory of mind' would capable of the first of these mappings but not the second.)

It is clear, moreover, that we can think of this further mapping in physical terms. For in learning to link sentences with our motives in this way, we perforce learn to map these sentences to the states of our brains -- say, to the invisible patterns of neural connectivity and activation -- which realize the motives. Just as a person who learns to eat thereby learns to fill her stomach, whether she knows about stomachs or not, so also a person who learns to describe motives by embedded sentences learns to link these sentences with her brain, whether she knows about brains or not. And since the sentences which we thus link with our brains are already mapped to the world, our mode of description also effects this further link between brain and world.

In learning this practice of description, therefore, we prereflecively and naturally link our repesenting mind/brains to the environment they represent. We thus in effect conceive the neural causes of behaviour in terms of the sententially described external situations which it is their adaptive function to produce (as in the case of desires, cf. Millikan 1984) or to reflect (as in the case of beliefs or thoughts). Hence we can regard the sentence-to-world mappings which inform our commonsense concept of motive as a way of thinking of the structures and activity which realize motivation in the brain.

To see this more clearly it may be useful to recall Patricia Churchland's (1986) description of the 'phase

space' conception of computational neuroscience. According to this view

...retinal neurons will specify a phase space, vestibular neurons will specify a phase space, and so forth. If the afferent system is to play a role in the organism's feeding, fleeing, and so forth, then afferent phase spaces will have to be co-ordinated with efferent phase spaces...to put it schematically , the first must be an "as-the-world-presents-itself" representation, whereas the second must be an "as-my-body-should-be" representation...tensors appear to be an efficient way in which the sensory representations can be transformed into output representations in motor phase spaces.

In light of the above we can see that commonsense psychology already uses sentences of natural language to specify the neural 'phase spaces' which Churchland indicates. Thus in taking it that a situation P is a cause of a perceptual experience that P and this of a belief that P, we use repetition of the sentences by which we describe perceptual input to signify both the proper functioning of the afferent phase-space and the correct formation of the 'as-the-world-presents-itself representation'. Again, in commonsense terms, the case in which a desire that P brings about the situation that P, and this a belief that P, is one in which the desire that P is pacified -- that is, caused to cease to govern action -- by belief in the situation which satisfies it. Here sentential repetition also describes the correct working of the efferent phase-space, and the 'as-my-body-should-be' representation. Finally, common sense performs the transformation of input into the output representations -- the job Churchland thinks may be done in the neural mode via tensors -- by logical transformations. The case in which a desire that P and a belief that if Q then P cause a desire that Q is one in which desire is transformed logically, so that the satisfaction of the derived desire and the truth of the belief which caused it suffice for the satisfaction of the original.

Our commonsense practice thus enables us to represent the working brain as a virtual semantic engine, that is, one whose inputs, internal workings, and outputs are specified via sentences and the situations they describe, and so in terms of the notions of truth, reason, and the satisfaction of desire. On an account of this kind there is no need for showing that commonsense mental states can be identified with neural states, for we are already taking commonsense description to relate (albeit indirectly) to the nervous system. And since the key aspect of commonsense psychological vocabulary we are considering is the use of embedded sentences of natural language, there is no question of reducing this vocabulary or of eliminating it in favour of another. Here as in many other cases we can see commonsense description as both the basis for science, and continuous with the further understanding which we hope science will provide.

The idea that commonsense psychological thinking indirectly represents neural activity in this way also seems a natural consequence of evolution. Roughly, we can see that an increase in the ability to understand and anticipate the behaviour of others should be an evolutionary advantage to members of a species who possess it, and the same holds for an increase in that ability determine how one is understood or anticipated by others. So we might expect that there would be circumstances in which evolution would cull and save in favour of both these abilities. Such a process seems to have been accelerated among the social primates, and particularly in our own species. (See, e.g., Deacon 1997.) So, it seems, over millenia we have evolved to possess interrelated abilities to use language and a conception of the mind. Hence these abilities unfold together, as in accord with innate structure, from early childhood; and they seem

aspects of a natural system by which we human beings relate to one another, that is, understand and influence the behaviour of members of our own species.

The behaviour-governing mechanisms of which we have thus gained some understanding and control, however, are realized in the brain and nervous system. This means both that they are normally out of sight, and that their causal and functional roles could not be rendered transparent to perception in any case (as shown by the fact that we with all our science are scarcely beginning to fathom them.) So, as we can say, the human species -- or the human nervous system -- has had to evolve the capacity to represent behaviour-governing states and events in nervous systems blindly, that is, without drawing on perceptual information about neurons as disposed in space and having features perceptibly related to the operations of computation and control which they perform. The sentential and semantic way of thinking of the causal role of neural states and events outlined above can be seen as a response to this problem. For it seems that this way of thinking has been precipitated by a process of building up a representation of the neural causes of behaviour as from the outside the body, that is, by mapping them to the external situations to which they are causally related.

4. Ways of thinking and identity.

These examples make it plausible to hold that our commonsense psychological concepts can be regarded as ways of thinking of the brain and nervous system. Just as in thinking of water we are thinking of H20, so in thinking of motives we are thinking of states and events in the brain. In both cases we think of these referents indirectly, that is, in a way which does not make the further nature of the referent explicit. In the case of motives, however, the indirect mode has the advantage of enabling us to process information about causal role which is indispensible for the co-ordination of our actions, and of which we have as yet no scientific specification.

We can also see that this gives rise to familiar kinds of problems of identity. In general, different concepts or ways of thinking represent their referents as having distinct properties; and if we fail to see that these properties are possessed by the same thing, we may think that we are dealing with two referents instead of one. Someone might think, for example, that water was a simple substance, and hence that it was distinct from H20, which is complex. Such an error would be intelligible, for in using the concept water we do not explicitly represent the referent as complex, as we do in using the concept H20.

Likewise, it seems, some have argued that desires could not be causes of action, on the grounds that causes and effects must be connected by empirical laws, whereas the connection between desires and the actions which satisfy them is semantic and a priori. This too is intelligible, for in relating desires and actions semantically we do not explicitly represent their relations as causal, as we might do if we developed a more scientific account of desire. But in this case there is also tension which is missing from the example of water and H20. For we have learned to regard semantic and causal relations as distinct, and this may make it hard to see that in the case of motives we use the one to encode information about the other.

5. Thinking of inner: the example of conceptual metaphor.

Let us now consider how we think of the mind as inner. No doubt, as mentioned above, our thinking in this way somehow reflects the fact that the events we describe in mental terms actually occur inside the body. (Hence we have the phrase 'in the mind or head' which seems, disconcertingly, to assimilate distinct sorts of internality.) But the physical location of these events by itself does nothing explain how we manage to represent them as internal, much less how we have come to represent them as internal (as well as phenomenal, subjective, and private) in the particular way we do. So how might we suppose that our particular mode of representation of internality has arisen?

Here it seems worth noting that we already make use of another mapping from the environment, by which we represent the internality of some mental events. Recently George Lakoff, Mark Turner, and a number of others have argued that we frequently represent via cognitive metaphor. In this we systematically map one domain of objects and properties (the source domain) to another (the target domain), and use the one to represent, or think about, the other (see Lakoff 1993). To take a relevant example, we often represent the mind in terms of the inside of a container, where this container can also be taken as the body. We can call this the metaphor of the mind/body container.

Metaphors from this family appear in many contexts, as when we say that someone who has failed to keep something concealed has spilled the beans, i.e. let things spill out of his mind/body container, and in a way that makes them difficult or impossible to replace. They are, however, particularly common in our conceptualisation of emotion (see, e.g. Koveces 1990). 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: the feelings of someone who is angry may seethe or simmer and so are agitated. 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 doesn't manage to let off steam, he may burst with anger, or explode with rage. We thus represent the spectrum of feeling between calmness and uncontrollable anger 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. By contrast a source of fear may give one 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. Here the opposition in the nature of feelings is marked by an oppositon in the properties of the metaphorical fluids to which we map them.

This is only one of very many examples of representation of the mind as an inner space or container (for some others see Hopkins 1999d); but it may indicate something of the tacit and systematic nature of such thinking.

6. Metaphoric representation and the concept of mind.

The kind of thinking evidenced in conceptual metaphor seems a basic way of effecting systematic comparisons, and one by which we can envisage the brain making use of existing neural prototypes in order to bring new domains into its representational scope. Let us therefore suppose that a capacity for this kind of thinking is also part of our evolutionary heritage, and consider how it might bear on

representation of the mind.

We said earlier that evolution apparently posed nervous systems the problem of coming to represent the behaviour-governing states of nervous systems without reliance on perceptual information bearing directly upon neural operation. In previous sections we speculated that in the case of belief, desire, and the like this achievement was facilitated by mapping the neural states in question to things which could be perceived, and hence to sentences, and via these to the environmental situations which the neural states were adapted to produce or reflect. The mapping of mental items to items in containers considered above seems a comparable form of representation. In both cases, as it seems, we represent the causal role of internal neural states or processes by mapping them to things external to the body. In this latter case, however, the mapping explicitly represents the innerness as well as the causal role of the processes to which it is applied. The emotions are represented as acting as a fluid might; and this activity takes place inside the mind/body container, as a fluid acts inside a vessel.

While the sentential mapping provides a powerful representation of the causal role of the items to which it is applied, it provides no image of their internality, nor of important internal aspects of their causal role. This is a lack which we can envisage as supplemented by the kind of mapping exemplified by the metaphor of the mind/body container. In this we represent events which are (i) perceptually and causally inscrutable and(ii) hidden inside the body by mapping them to others which are (i) perceptually intelligible but which may be (ii) hidden in containers in the external environment. We thus use our information about containment in space in the external environment to create an image of containment in an inner space, which we apply to neural events which we characterize as both mental and having a significant internal role.

7. Wittgenstein and the metaphoric origin of the problems of mind.

This approach via metaphor can be compared with Wittgenstein's (1954) discussion of these issues. As is familiar, Wittgenstein held that philosophical problems arise from unconscious comparisons between the way we use words in different regions of language; and he applied this with particular effect to the problems of mind. These, he said, arose from a 'picture' of the mind:

425. In numberless cases we exert ourselves to find a picture and once it is found the application as it were comes about of itself. In this case we already have a picture which forces itself upon us at every turn, -- but does not help us out of the difficulty, which only begins here...
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.

This picture, as Wittgenstein describes it, is closely linked with the dualistic picture of the mind as sketched above. In it the mind is conceived as an enclosed space, whose contents are detected by a process analogous to sight (introspection). This space is also distinct from that inside the body, for its

contents are psychological as opposed to physical ('one is not thinking of brain-processes, but of thought-processes.) Wittgenstein observed that on this picture of the mind the phenomenal properties of experience would be unknowable to anyone besides the subject, and that this in turn would render these properties incapable of being referred to or described in a language whose use was determined by public criteria. (The inference is immediate, since something which plays no role in determining the use of a term is not referred to or described by it.) Since these properties were supposed to described by such phrases of public languages as 'the feeling of pain', he took this as a partial reductio of the view, which he put as follows:

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.

This argument clearly turns on an instance of the metaphor of the mind/body container, that is, the box which here represents the mind. So, as we can say, Wittgenstein took the problem of other minds to be partly constitued by the imposition of this metaphor. He seems to have taken the same metaphor to give rise to the problem of consciousness; but in this case he did not make the connection fully explicit. For example he writes:

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

These remarks are highly compressed, but it seems that we can take Wittgenstein's thinking to run as follows. 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. (The box in the previous remark is replaced by the pot in this.) In order to understand these internal events we form a representation (picture) of them, in terms of the mind/body container (the picture of the pot). But in using this picture, we perforce represent the internal events as occurring, not in the physical space of the body where they actually occur, but rather as in a space of a distinct kind. We are thus misled by our form of representation, in such a way that we think of the internal events with which we are concerned as

occurring not in the pot (the body), but in the picture of the pot (the metaphorical space internal to the mind/body container).

8. The metaphor of containment and the notion of a virtual inner space.

We noted above that there is reason to hold that metaphorical thinking is an important part of our conceptual repertoire, and that we make particular use of it in representing the internality of the mind. Wittgenstein's discussion thus represents these problems as springing from a preconscious mode of thought to which we have access through other disciplines, and which we have other grounds for considering as significant.

We can make Wittgenstein's claims here clearer by using the notion of a virtual space, which may contain virtual objects or properties. A virtual entity -- as in the example of the mind construed as a virtual semantic engine above -- is one which is not real, but is rather an artefact of our modes of representation. Thus we can think of the space shown in a mirror, not as that of the actual room, including ourselves, which is reflected in it, but as an alternative space which we might enter by passing through the glass. Taken this way the space through the looking-glass is a virtual space -- a space we can consider in detail, think of ourselves as entering and having adventures in, etc., but which nonetheless is not real.

The same holds for things or properties reflected in the mirror; these can be construed as virtual entities, which we would be able to touch only if we passed through the glass. This also illustrates how a representational change from real to virtual entities may leave the properties of the entities in question apparently unchanged. (Of course the looking-glass world may be very different in nature from this one -- it may be one in which objects present no resistance to the hands for example. In this case the switch to virtual entities will be taken to affect the properties which entities have.)

We can treat the space and items presented by a cinema or television screen, or that shown in a video game, in the same way; and if the representation with which we are dealing is part of a sufficiently comprehensive or compelling illusion, we may characterize it and the other virtual items in it as a virtual reality. Thus we create a virtual space when we in one way or another treat a space which is in fact only represented or imagined as if it were real; and the limiting case of this, as I am construing the notion, is believing that a virtual space is real.

Believing that a virtual space is real is another form of believing that there are more things (in this case more spaces and entities in them) than there actually are. This is the idea involved in Wittgenstein's metaphor. Someone who insisted that there had to be something boiling in a picture of a pot would be taking a represented space, represented as having something boiling in it (the space represented as inside the picture of the boiling pot) as a real space with someone boiling in it. This would be creating a virtual space, and a virtual boiling liquid, in the sense spelled out above. (The space in the pictured pot and the liquid in it would be like the space and objects behind the looking glass, taken as real.) So the idea which we are to consider is that the apparently distinct inner space in which we are inclined to insist that conscious events occur, and the properties manifest in that space, are virtual entities, and therefore artefacts of our representation of the mind as a container.

For an example of the use of this concepion we may take the 'phenomenal cosmology' introduced by Metzinger above, with its range of 'individual spaces', which constitute 'subjective universes', which appear with human birth and disappear with human death. The view under consideration is that this picture is not literally true. It is not actually the case that the physical space in the vicinity of persons or other conscious creatures is holed with other spaces of a mysteriously different kind containing experiences. Rather on this account the spaces Metzinger is talking about -- and those depicted in the figures we have drawn above -- are virtual spaces, spaces which we are inclined to take as real because of the way we represent things in spatial terms.

9. Virtual inner space and the representation of real inner events.

This brings us to a second aspect of Wittgenstein's remark, which is that the metaphor of the mind as a container actually serves as a representation of events in the body. (The picture of the boiling pot, although it does not actually contain anything boiling, does nonetheless represent boiling as going on in the pot.) We can see this clearly in the connected case of thinking of anger as a hot fluid inside us, or representing emotion in terms of internal fluids generally. This way of thinking serves as a representation of things that go on in us when we experience emotion, and hence it serves as a representation of physical processes which are otherwise unknown to us, but which are in fact contained within the body.

In speaking of levels of temperature of an imaginary emotion-fluid, for example, we seem to be indexing behaviour-governing neurophysiological processes, particularly those which are occurring in the autonomic nervous system. Since this system produces and monitors emotion-related changes in the guts, for example, it seems that we are also mapping its activity in feeling that our guts have run cold, or turned to water, or whatever. How such mappings work in detail we have yet to discover, but it seems reasonable to suppose that they underpin the metaphor, so that its targets are ultimately physical.

As a first approximation, therefore, we can see the metaphor of the mind as a container as representing something inner by something inner: we can take the representation of the mind as an inner space as a primitive natural way of representing the working of the nervous system in the space inside the body. Like the semantic representation of causal role discussed above, this way of thinking seems designed to work in the absence of explicit knowledge about its representational function. Hence the question whether we are dealing with representations of the real space and events inside our bodies, or with an alternative virtual space containing virtual events, will depend upon how we construe the representations in question. It seems that our natural construction -- one that we have to work to overcome -- is in terms of the virtual space and entities of the Cartesian tradition.

10. Virtual inner space and the apparently non-physical features of consciousness.

This approach enables us to use our understanding of metaphoric representation to the features of mind which we regard as problematic. Such thinking, as we have seen, maps objects, events, and properties across domains. It is therefore a requirement for such thinking that the source and target domains be both

appropriately connected to, and appropriately differentiated from, one another. If sources are not appropriately connected with their targets they cannot be used to represent them; and if they are connected but not differentiated from their targets the one may be confused with the other. In both cases -- that in which the source of a would-be metaphor is not linked with a target, and that in which a source is confused with a target -- metaphoric thinking will fail.

Lakoff represents the maintenance of correct source/target relations in terms of what he calls the Invariance Principle. This principle, as he says,

...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 target domain structure automatically limits what can be mapped. (1993, p 215-6)

To think that someone given a kick would possess that kick afterwards would be to fail properly to distinguish the source domain of physical objects transfered from one person to another in givings, from the target domain of kickings of one person by another. According to much psychological and psychiatric observation, such failures of source-target relations are shown in a number of psychological disturbances characterized by concrete thinking. To take a simple example, a ten-year-old autistic girl showed catastrophic anxiety when a nurse, about to do a blood test, said 'Give me your hand; it won't hurt.' She calmed down immediately when another person said 'stretch out your index finger'. She had apparently understood the situation -- in violation of the invariance principle -- as one in which if she gave her hand, she would not have it afterwards.

We tacitly maintain a complex set of such source-target relations when we think of the mind as a container. For example if we think of anger as a hot fluid inside us, and so actually feel the anger in this way, we still do not think that if someone's anger wells up, boils over, or spills out, this anger will subsequently be found spattered on the carpet. To use the metaphor thus would clearly be to think of anger and its locus in too concrete a way, and most people automatically do not do so. Rather we subtly and systematically de-concretize and so de-*physicalize* both the space occupied by the anger-as-fluid and the metaphoric fluid itself.

This means that we tacitly treat the anger-space as a non-physical space, not to be confused with the actual internal space with which, nonetheless, it may phenomenologically overlap; and likewise we treat the anger-fluid as a non-physical fluid, not to be confused with physical things actually inside us. To say this, however, is to say that the anger-fluid and the anger-space are themselves virtual entities, which we represent as differing from the real ones on which they are modelled. Still we can see that this representational de-physicalization ultimately involves nothing which is really non-physical. Rather it flows from the tacit imposition of the requirement of avoiding concrete thinking (something like the invariance principle) upon a mapping which has both physical sources (physical fluids and containers) and physical targets (changes inside the body involved in emotion). Since nothing which is both real and non-physical actually comes into question, we can say that the apparent non-physicality of the anger-space and fluid are a sort of cognitive illusion, engendered by this spatial mode of representing the inner. So it seems that the same process might likewise account for the apparent non-physicality of the inner

space and contents involved in our everyday conception of the mind.

If this approach is correct, then it may be possible to see the features of experience which constitute the problem of consciousness as virtual features, that is, as appearances engendered by our way of representing the neural events which realize experience. This would be a substantive account of the way our modes of presentation generate the problem of consciousness.

On this account the problematic distinction between the inner and the outer is produced by our representing experience as occurring in one or another kind of inner field or space -- visual space, auditory space, the space in which we feel pain, etc. -- where these 'spaces' are modelled on space as we perceive it. (This is the part of role of the container metaphor.) This means, in effect, that we do not directly represent the neural events which realize experience as in the bodily space in which they actually occurr, but rather make use of an image of space derived from the perception of space outside the body. This second-hand image is therefore liable to be construed as that of a virtual space.

We can start to see this in the dualistic image of visual experience, as portrayed in Figure 2. In this case we seem to be representing neural events in the visual system by a straightforward mapping from outer to inner -- a metaphoric internalization of the space and process of visual perception itself. It is as if in this way of thinking of experience the ordinary space in which we see things had simply been transposed inside (and hence stripped of its physical substance and made the object of a further special sense) so as to become the non-physical, quasi-spatial visual field, which we think of as somehow internal.

If we were to think of this mapping from outside to inside simply as a mode of representation of what goes on inside the body, comparable to the mapping effected by our sentential specification of desires and beliefs, it would present no conceptual or metaphysical difficulty. But insofar as we construe it in terms of an alternative space, somehow superimposed upon that inside the body, we must perforce also feel that the entities and properties displayed in that space are real but not those of concrete physical things. This change is comparableto that involved in thinking of reflected properties as existing, not in real space, but rather in the space accessible through the looking glass. In this case, however, the imposition of an alternative space requires us to construe these properties as phenomenal in a sense which is opposed to the physical.

Again, if we represent something as phenomenal just insofar as we represent it as perceived within a virtual internal space, then there will be no more to phenomenal objects than is manifested in the space in which they appear, so that what is phenomenal will also be subjective. So this approach may also go some way towards explaining why we feel that the esse of phenomenal properties is their percipi, and why these properties seem, puzzlingly, both distinct and yet not distinct from our apprehension of them.

Finally, since these virtual properties are represented as perceived within the space representing a single mind, they also seem private. Here again, if this account is correct, the representation of the mind as a container portrays something which is real. For it is true that our experiences occur within us, and hence that each of us is related to his or her own experiences, and knows of them, in a way which would be impossible for anyone else. In the metaphoric mode of representation we are considering, however, this is

registered in terms of containment in space. The impossibility that another should be cognitively related as I am to events in my own nervous system is thus represented as the impossibility that another should see into my inner space.

Clearly these hypotheses require to be developed further. The overall idea, however, is straightforward. It seems that the mind is simply presented to us, but this is not so. Rather our apprehension of the mental is doubly mediated; the manifest image of the mind is constructed by mappings from the prior manifest image of the physical world. These mappings enable us to regulate and co-ordinate our behaviour, by representing the inner causes of behaviour in terms of a virtual semantic engine (the sentential mapping) housed in a virtual internal space (the container mapping). These ways of thinking are rooted in, and embed, prior ways of thinking of the physical world; but they admit of misunderstanding, which seems to set the mental apart from the physical. Thus, remarkably, the way in which we represent the causal role of desire can seem to place desire outside the causal realm, and the way we represent mental events as occurring within our bodies can seem to place them outside physical space.

The explanatory gap between the physical and the phenomenal thus has the same origin as that between causes and reasons. In both cases a structural difference in our ways of thinking of the mental as opposed to the physical creates the impression of a deep difference in the things represented. If the argument here is correct the appearance in both cases is illusory. For apart from these differences in our ways of representing things there is no real distinction in nature to which these profound apparent differences correspond.