A partir dessas formulações e frente ao insucesso do paradigma social, material, sem levar em conta todas as circunstâncias relevantes, o paradigma liberal, formal, encontra razões epistemológicas para abandonar a complexidade à sua própria autoregulamentação. Frente aos dois, o paradigma procedimental pode tentar abarcar o âmbito da complexidade das questões relevantes para o tratamento da desigualdade, em que os próprios destinatários do direito, como seus autores, podem corrigir os rumos dos acontecimentos, num processo de aprendizagem falível:

> Todo aquele que tenta enfrentar as perspectivas reformistas, servindo-se apenas dos argumentos triviais que destacam a complexidade, confunde legitimidade com eficiência e desconhece o fato de que as instituições do Estado de direito não visam simplesmente reduzir a complexidade, mas procuram mantê-la através de uma contra-regulação, a fim de estabilizar a tensão entre facticidade e validade.[67]

Trata-se de entender a constituição e, portanto, o direito "como sendo a instituição de um processo de aprendizagem falível, através do qual a sociedade vence, passo a passo, sua natural incapacidade para uma autotematização normativa".[68] Como visto, as diferenciações no direito processual, como parte do paradigma dos direitos sociais, bem como a questão do feminismo, são os exemplos aportados para indicar de que modo se evitaria o paternalismo tendencialmente ligado a esse paradigma, permitindo que a liberdade e a igualdade sejam melhor realizadas do que no paradigma liberal.

67 TrFG2, p. 188 [FG, p. 535].
68 TrFG2, p. 189 [FG, p. 535-536].

Kriterion vol.1 no.se Belo Horizonte 2006

A QUASI-MATERIALIST,QUASI-DUALIST SOLUTION TO THE MIND-BODY PROBLEM

*John-Michael Kuczynski**

RESUMO *Se o mental pode afetar ou ser afetado pelo físico, então o mental deve ser ele mesmo físico. Se não fosse assim, as explicações do mundo físico não poderiam ser fechadas — e elas são fechadas. Há razões para se pensar que o materialismo é falso, tanto em suas versões redutivistas quanto nas não redutivistas. Mas como explicar então a aparente sensibilidade do físico ao mental e do mental ao físico? A única solução possível parece ser a seguinte: objetos físicos são na realidade projeções ou isomorfos de objetos cujas propriedades essenciais são mentais. Um modo um pouco menos preciso de apresentar essa tese é o de dizer que propriedades constitutivas, i.e. não estruturais e não fenomenais, de objetos físicos são mentais, i.e. são propriedades tais as que habitualmente encontramos apenas por "introspecção". A cadeira, na medida em que a conheço através da percepção sensorial e de hipóteses estritamente baseadas na percepção, é um tipo de sombra de um objeto que é exatamente como ela, com a única diferença de suas propriedades essenciais serem mentais. Esse raciocínio, embora radicalmente contra-intuitivo, explica a aparente sensibilidade do mental ao físico e inversamente, sem se expor às críticas feitas ao materialismo, ao interacionismo dualista e ao epifenomenalismo.*

Palavras-chave *Filosofia da mente, Problema do Corpo-Mente, Fisicalismo, Dualismo*

* Professor do Departamento de Filosofia da University of California, Santa-Bárbara.
Artigo recebido em set./2003 e aprovado em mar./2004.

ABSTRACT *If the mental can affect, or be affected by, the physical, then the mental must itself be physical. Otherwise the physical world would not be explanatorily closed. But it is closed. There are reasons to hold that materialism (in both its reductive and non-reductive varieties) is false. So how are we to explain the apparent responsiveness of the physical to the mental and vice versa? The only possible solution seems to be this: physical objects are really projections or isomorphs of objects whose essential properties are mental. (A slightly less accurate way of putting this would be to say: the constitutive — i.e. the non-structural and non-phenomenal — properties of physical objects are mental, i.e. are such as we are used to encountering only in "introspection".) The chair, qua thing that I can know through sense perception, and through hypotheses based strictly thereupon, is a kind of shadow of an object that is exactly like it, except that this other objects essential properties are mental. This line of thought, though radically counterintuitive, explains the apparent responsiveness of the mental to the physical, and vice versa, without being open to any of the criticisms to which materialism, dualistic interaction ism, and epiphenomenalism are open.*

Key-words *Philosoph of the mind, Problem of the Body-Mind, Physicalism, Dualism*

I. The Scope and Methodology of the Present Paper

The mental and the physical are *causally* so well integrated with each other that, it would seem, they *must* be identical. To be more precise, given that mind and matter are causally responsive to each other, and *given also that the physical world is causally closed — i.e. given that the cause of any physical event is another physical event* — it follows that mind is a kind of matter.

At the same time, nothing in the physical world — in the brain, in particular— seems to 'disclose' mentality. When you look at a brain, you see beige tissue (or cells or molecules — depending on how you are looking at the brain): you *don't* see ideas, feelings, intentions; and you *don't* see anything that needs to be *explained* in terms of ideas, feelings, intentions etc. What you see is, by all accounts, no more in need of *mentalistic* explanation than the behaviour of a brick)

No matter how thoroughly you studied a brain, you would never, in the course of those studies, encounter an idea or a feeling. You would encounter cells, molecules, atoms, and so forth: but never a thought or a desire.

It is often said that our grounds for identifying mental phenomena — e.g. pains, beliefs, feelings — with brain-states are perfectly comparable to our grounds for identifying water with H_2O or heat with molecular motion or light with streams of photons. But this simply isn't true. This becomes evident when we attend to the close connection between the concept of physicality and that of perceptibility.

Molecules cannot be seen — our technology doesn't currently permit it.[2] And streams of photons cannot *possibly* be seen — the laws of physics do not permit it.[3] But we could, in principle, create an object that was, structurally, just like a water molecule but was trillions of times larger: so that we *could* actually see it. This model would graphically display the very objects and properties that we believe, on theoretical grounds, to constitute water. The same is true *mutatis mutandis of* molecular motion and streams of photons. The explanatorily relevant features of these entities could, in principle, be given a graphic or plastic representation. We could construct a visual model that gave tangible expression to the features of these entities that were theoretically important — that are implicated in the theories that posit those entities.

To put it very roughly, if we were small enough, or if the aforementioned theoretical entities were big enough, we could *see* molecular motion: we could *see* the theoretical entities implicated in identifications like *water is* H_2O and *heat is molecular motion.*

(The identification of light with a stream of photons poses special problems: we couldn't possibly *see* photons. That is why, in the previous paragraph, I spoke of seeing models of photons that displayed the theoretically relevant features of them. But the basic idea prevails: although we couldn't see photons, we *could* see things that modelled the theoretically significant features of photons — the features ascribed to them in the theories that posit their existence.)

But nothing even remotely comparable is possible in the case *of* an idea or a thought or a feeling. It is true that we identify these things with brain-states. But given any constituent *of* a brain-state — any aggregate of cells, any individual cell, any molecule, any atom — if we created a physical model of

1 See Eccles [1].

2 This isn't quite true — it is subject to some delicate qualifications. But these aren't really of importance in the present context.

3 What follows was anticipated by a remark that Leibniz makes in the *Monadology.* He says that if we could walk around inside a brain, we would never see thoughts, feelings, desires, and so on; only various physiological processes.

that constituent that was large enough for us to see it, that model would to no degree whatsoever exhibit any of the properties characteristic of ideas, thoughts or feelings. When we identify a brain-state with, say, a perception, we are identifying a brain-state with something that necessarily has the properties of being representational, of having a felt-quality, of purposiveness, of having a kind of subjectivity, and so forth. But if you were small enough to *see* the cellular or molecular activity with which, supposedly, the perception is identical, you would not see any of these properties: you would see things that no more disclosed the properties of being representational, of having a felt quality, and so forth, than a rock. Compare: if you were small enough to *see* H_2O molecules, you *would* be able to see the properties of those things that are ascribed to them in the theory that identifies them with water.

There is a close connection between the concept of physicality and that of perceptibility. (Later on, we will see just how close this connection is.) To be sure, not everything physical is perceptible. In some cases, it is technologically impossible to see a physical entity. **In** other cases, it is *absolutely* impossible (given the role that photons play in visual perception, it would be *absolutely* impossible — causally impossible and, I think, conceptually impossible — to see photons.) But even though some physical entities cannot be seen, representations of their explanatorily or theoretically important features *can* always be seen: we have no trouble creating a perceptible model of the theoretically important features of H_2O molecules. But there is absolutely *no* prospect of creating a perceptible model of the distinguishing characteristics of mental entities: representationality, deliberateness, phenomenology ('felt quality'), subjectivity, and so forth.

So it is simply not true that our grounds for identifying (say) pain with c-fibre stimulation are comparable to our grounds for identifying water with H_2O. When we identify water with H_2O, we are, in effect, claiming that, if we were small enough, or water was large enough, we would *see* H_2O molecules when we walked about in a body of water. (A corresponding claim is true of the identification of light with photons, though some additional complications are involved in this case.) But when we claim that (say) pain is c-fibre stimulation, we cannot, if we are sane, be claiming that, if we were small enough (or brains were big enough), we would *see* pain as we walked about the interior of a brain. And, of course, the same is true of the identification of any mental entity with a brain-state. So the logic behind identifications like *pain is c -fibre stimulation* is dramatically different from that behind identifications like *water is* $\boldsymbol{H_2O}$.

So we are in between a rock and a hard place. If we accept dualism — the

view that mind and body are distinct — it becomes hard to account for the obvious *causal* integratedness of the mental and the physical. But if we accept materialism — the view that mind and body are one — we apparently destroy the *explanatory* unity characteristic, if not *definitive*, of the physical domain: we introduce into the physical world something which resists physical explanation, something which couldn't conceivably be encountered in the physical world — even in the indirect sense, described above, in which photons can be encountered — and whose physicality is *ipso facto* open to question.

My purpose in this paper is to thread a path through the Scylla of dualism and the Charybdis of materialism. The doctrine I set forth will seem to many to be a kind of materialism, and to others it will seem to be a kind of dualism. My suspicion is that most would regard it as *a* form — albeit and unusual one — of materialism. My own view is that, strictly speaking, my position is more correctly described as a form of dualism.

The solution set forth here is not wholly new. Its point of departure lies in a comment made by Russell:

> I conclude that, while mental events and their qualities can be known without inference, physical events are known only as regards their space-time *structure.* [My emphasis.] The qualities that compose such events are unknown — so completely unknown that we cannot say either that they are, or that they are not, different from the qualities that we know as belonging to mental events.[4]

Ultimately, all I do in this paper is to take Russell's remark seriously. I suggest that, first of all, we *suppose* that the 'qualities that compose [physical] events' *are* mental. I maintain that if we make this supposition, we can (i) account for the causal integratedness of the mental and the physical; and we can (ii) account for the difficulty we have explaining the existence of mind in terms of matter.

It might seem that, in taking this view — in taking the view that the 'qualities that compose [physical entities]' are mental — I am espousing a form of *materialism.* ('Surely if one says that the physical is *composed* of such and such, then such and such is ipso facto physical, whatever else such and such may be.')

My response to this as follows. The concept of physicality is *a structural*

4 Russell [1] p. 247. Elsewhere Russell writes:
'[W]e have found it necessary to emphasize the extremely abstract character of physical knowledge, and that fact that physics leaves open all kinds of possibilities as to the intrinsic character of the world to which its equations apply. There is nothing in physics to prove that the physical world is radically different in character from the mental world...The only legitimate attitude about the physical world seems to be one of complete agnosticism as regards all but its mathematical properties.' Russell [2] pp. 270-271.

concept; when we ascribe physicality to something, we are saying that it has structural or formal properties of a certain kind. (This view is counter-intuitive; but, I believe, capable of a cogent defence.) Now objects cannot have *only* structural properties; such properties must be 'fleshed out' somehow. The properties that 'flesh out' or 'embody' a structure cannot *themselves* always be purely structural; to deny this would involve some kind of a vicious regress. So those properties are non-structural and therefore non-physical. I will make heavy use of Kant's distinction between 'phenomena' and 'noumena' (a distinction that, I think, may be implicit in Russell's remark).

More specifically, I will maintain that, at least where some physical entities are concerned, those entities may regarded as the 'phenomena' (in Kant's sense) whose corresponding 'noumena' are mental. The physical entities in question would be brain-events and states: so the 'qualities that compose such events [and states]' are, I will maintain, mental in nature.

II. Is Dualism Compatible with Interactionism?

Mental events seem to be responsive to physical events and *vice versa*. *A* hot iron is pressed to my skin, and I feel pain: here a mental event seems to occur *in response to* a physical event. I see a rabid dog running towards me, and I subsequently bolt in terror: here a physical event seem to occur *in response to* a certain mental state.

How are we to account for the apparent responsiveness of the mental to the physical and *vice versa?* The most obvious answer is this: the mental and the physical do not just *seem* to be responsive to each other: they really *are* responsive to each other; they really do interact. Is this answer tenable?

Before we can answer this question, we must note one thing: if the physical and the mental interact, then the mental must itself be physical. Why is this? Suppose that, indeed, mind and matter do interact; and suppose, further, that mental events were *not* themselves physical. In that case, entities that did not themselves fall within the scope of physical laws could affect entities that *did* fall within the scope of such laws. (If mental entities are not physical, then of course mental entities do not fall within the scope of physical laws.) Every time something that falls within the scope of a physical law is affected by something that does not fall within the scope of such a law, an exception to that law is thereby generated. For in such a case, the behaviour of the affected entity *itself* falls outside the scope of that law, as its behaviour is now a function of the behaviour of some entity that doesn't fall within the scope of that law (namely, some mental entity). In that case, the physical world would not

be *explanatorily* closed: in order to explain physical events, it would be necessary to take mental events into account.

But the physical world *is* explanatorily self-contained. In any case, all the available empirical data supports this. To explain the movements of my body, it is not in principle necessary to take my mental states into account. By all accounts, my body no more falls outside the scope of physical law and, therefore, of physical explanation than do rocks and billiard balls. (Of course, it is *easier* to predict someone's physical behaviour by taking his mental states into account. But that is irrelevant. It is easier to predict the behaviour of a computer by thinking of it as doing sums. But that doesn't mean that the computer's behaviour falls outside the scope of physical law or explanation.) So either the mental doesn't affect the physical or the mental is itself physical.'

This argument can be put another way. If mental events could affect physical events, and mental events were not themselves physical, then given any alleged law of physical nature, some mental event could intercede in the course of physical events and generate an exception to that 'law'. But physical laws do not admit of exceptions. (If a true exception is found to some physical 'law', then it is *ipso facto* not a law.) So if the mental could affect the physical, and the mental were not itself physical, then there would be no laws of physical nature. But there are such laws.

So either the mental doesn't affect the physical or the mental is itself physical. So interactionism is true only if materialism is also true.

> But [asks an imaginary interlocutor] mightn't a limited kind of interactionism be compatible with dualism? Suppose that (i) physical phenomena could affect mental phenomena, but (ii) mental phenomena could not, in their turn, affect physical phenomena. Under these circumstances, if mental phenomena were non-physical, this fact would not entail either that there were no physical laws or (what may be just a different way of phrasing the same point) that the physical world was not explanatorily self-contained. After all, full-blown, bidirectional interactionism (the doctrine that matter affects mind and vice versa) is incompatible with dualism because, if mind affects matter and mind does not itself fall within the scope of physical law, then the physical world is explanatorily open. But, it seems, if matter affects mind, but not vice versa, the physical world would remain explanatorily self-contained: so maybe we can reconcile dualism with a limited form of interactionism.[6]

This view set forth by the objector — that matter affects mind but not *vice versa* — is known as *epiphenomenalism*. Epiphenomenalism is not tenable; dualism is not compatible with matter's being able to affect mind. Why is this?

5 See Kim [1].
6 This is what David Chalmers holds. See Chalmers [1].

Causation is bi-directional: roughly, x affects y just in case y affects x. (This is subject to a qualification that we will get to in a moment.) I cannot move the rock without the rock's affecting me in some way. From a purely *physical* standpoint, the rock is no more passive with respect to me when I move it than I am with respect to it. It is only from a pragmatic or psychological standpoint that, in such a transaction, I can be said to be more 'active' than the rock. Activity and passivity are concepts that apply to the human, not the physical, world.

So if physical events bring about mental events, then mental events bring about physical events. And, as we have already seen, if mental events can affect physical events, while not themselves being physical, then the physical world is explanatorily open. But it isn't explanatorily open. So dualism is incompatible with *any* kind of interactionism, no matter how limited.

> You said that if x affects y, then y affects x. That is plainly false. If I break a window by throwing a brick at it, I affect the window but the window doesn't affect me. So causation is not bi-directional. So the argument you just gave isn't sound.

If x's affecting y is mediated through a series of intervening events, then x may affect y without y's affecting x in its turn. But if x affects *y directly* this is not possible. I can throw a rock at a window without being affected by the shattering of the window. But my hand cannot affect the window without being affected by the rock in its turn.

III. Is Some Form of Materialism True?

So there is an excellent reason to hold that some form of materialism is correct. Matter seems to affect mind and *vice versa.* Unless materialism is true, mind cannot affect matter and matter cannot affect mind. So it seems practically incontrovertible that materialism is correct. But we will soon see that matters are not quite so straightforward.

There are countless forms of materialism. But *ultimately* — if we ignore sub-categorial differences — there are but two varieties. I will refer to these as *reductive* and *non-reductive* materialism. What is reductive materialism and what is non-reductive materialism?

Reductive materialism is the view that first-person entities are really identical with, or at least constituted by, *paradigmatically physical* entities. Something is paradigmatically physical if it falls within the scope of the so-called 'physical sciences' — physics, chemistry, biology. Examples of paradigmatically physical entities are atoms, molecules, cells, organs, planets, and the

forces that govern their interrelations. It is not easy to say in virtue of what, precisely, something falls within the scope of these sciences. (Later on we will come up with an answer.) But, at an intuitive level, the meaning of the expression 'paradigmatically physical entity' should still be clear. Roughly, some-thing is paradigmatically physical if there cannot be any serious question as to whether to classify it as physical. There can be no serious question as to whether to classify a rock, an atom, or a cloud as physical. But there can be such a question in connection with e.g. a belief. (This formulation will be refined shortly.)

According to reductive materialism, pains, tickles, beliefs etc. are physical *solely* because they are identical with, or constituted by, paradigmatically physical entities — with neural events, brain-states, displacements of certain kinds molecules or atoms.

Non-reductive materialism is the view that first-person entities are *not* identical with paradigmatically physical entities but are physical anyway. Pain is not identical with, or constituted by, c-fibre stimulation or any other paradigmatically physical entity. Pain is what it seems to be, and nothing else: it isn't secretly identical with e.g. c-fibre stimulation. But [says the non-reductive materialist] pain is still physical. Searle holds this view.'

Non-reductive Materialism

First let us consider non-reductive materialism. The following argument casts serious doubt on the validity of this thesis.

The laws discovered by the physical sciences govern *paradigmatically* physical entities and paradigmatically physical entities *alone.* The laws of physics do not concern headaches and tickles — unless, of course, headaches and tickles are really paradigmatically physical entities in disguise. But according to non-reductive materialism, that is specifically what they are not.

An object that couldn't interact with any paradigmatically physical entity surely wouldn't itself qualify as physical. Imagine a 'physical' object that had absolutely no effect on atoms, molecules, rocks, trees, retinae, nerve-endings — that was just a kind of impotent phantom. Such a thing, indeed, would be totally undetectable; for a thing is detectable only if it has effects on our bodies, which are paradigmatically physical. And that thing, by supposition, would have no effects on anything paradigmatically physical — it wouldn't affect

7 J.J. Smart was a reductive materialist. See Smart [1].
8 Searle [1] p. 49.

atoms, molecules, and so forth. It is very hard to see how such entity — being totally undetectable and, indeed, without any effect on anything paradigmatic-ally physical — would possibly qualify as physical.

So if mental entities are physical, they must be capable of affecting paradigmatically physical entities. So any materialist, whether reductive or not, is committed to holding that mental entities affect physical entities. At the same time, the non-reductive materialist says that mental entities are not identical with *paradigmatically* physical entities — are not identical with atoms or molecules or the things composed thereof. As we noted, physical laws govern *paradigmatically* physical entities only (the govern headaches and tickles *only* if these things are identical with paradigmatically physical entities). So if the non-reductive materialist is right, then entities that didn't themselves within the scope of the laws of physics could affect entities that did fall within the scope of such laws. This would mean, as we saw earlier, that there would be no laws of physics and that the paradigmatically physical world would not be explanatorily self-contained. But it is self-contained, and there are laws of physics. So non-reductive materialism is inconsistent with the fact that physical world is causally and explanatorily self-contained.

Non-reductive materialism is, I suggest, just Cartesian dualism in disguise. Cartesian dualism is a dualism of the mental and the physical. Non-reductive materialism is a dualism of the paradigmatically physical and the non-paradigmatically physical. But the term 'non-paradigmatically physical' covers just what Descartes call the 'mental', and the term 'paradigmatically physical' covers just what Descartes call the 'physical'. So non-reductive materialism is Cartesian dualism under the cloak of a new terminology; and it is therefore just as incapable of explaining the apparent responsiveness of the mental to the physical as is Cartesian dualism.

From here on out, whenever I use the term 'physical' to refer only to *paradigmatically* physical entities, and the term 'materialism' to refer to *reductive materialism*. This is justified by the fact that non-reductive 'materialism' really isn't materialism at all and that anything that isn't paradigmatic-ally physical isn't physical at all.

An Argument Against Reductive Materialism

So if any form materialism is correct, it is *reductive* materialism. In this section, I will outline an argument to the effect that reductive materialism is false.

Earlier I defined reductive materialism as the view that mental entities are identical with, or *constituted by*, paradigmatically physical entities. First of all, what is the difference between constitution and identity? Imagine a figurine that is made of clay. Is that figurine *identical* with the clay of which it is made? Well, you could destroy the figurine without destroying the clay. So the clay and the figure have different properties. Hence, the figurine is not *identical* with the clay. But every fact about the statue — whether it is beautiful, how much it weighs, etc. — is obviously fixed by some fact about the clay (e.g. the aesthetic properties of the statue are fixed by the shape that the clay has at a given time); and this, of course, is because the statue, while not *identical* with the clay, is *made up* of it — is, as we say, *constituted* by it.[9]

The distinction between constitution and identity is of some importance in connection with the mind-body problem. Reductive materialists hold that spatio-temporal world is *constituted* by interactions among elementary physical particles — quarks, muons, mesons, and so on. (Henceforth, we will refer to such interactions as *atomic interactions*, even though, technically, they should be called 'sub-atomic interactions'.) But strictly speaking it is not widely held that all physical entities are *identical* with sets of atomic interactions. My heart right now is *constituted* by certain atomic interactions. But my heart isn't identical with these interactions. For in a moment it will be constituted by completely different interactions. In a few years it will be composed of completely new particles altogether. My heart can, and will, survive the extinction of *this or that particular set of interactions. So* my heart has different 'modal properties' from the atomic interactions which currently constitute it, and therefore isn't identical with them.

For a certain physical state of affairs to obtain — for there to be a set of atomic interactions with certain properties — is really just for certain physical predicates or, equivalently, physical *concepts* to be instantiated in a certain region of space-time. (I am using the term 'concept' as a rough synonym for 'predicate'; I am *not* using the term 'concept' to denote anything *mental.* I will elucidate this qualification in a moment.) For a particle with mass x and charge y to be moving with velocity z in space-time region R is just for the concept *particle with mass x and charge y to be moving with velocity z* to be instantiated in R.

Reductive materialism must obviously hold that mental phenomena are in space. For reductive materialism holds that mental phenomena are identical with physical phenomena, and obviously physical phenomena are in space. Supposing that reductive materialism is right about this, what would it be for a

9 This argument is due, I think, to Bernard Wiggins.

certain mental event to occur in space-time region R? For such an event to occur in space-time region R would simply be for a certain mental *concept* to be instantiated in R. For there to be a surge of anger in R is just for the concept *surge of anger* to be instantiated in R. (I am not myself saying that mental entities are in space. I am saying that *if* as materialism holds, mental entities are in space, then for such and such a mental event of to occur in region R just is for such and such a mental concept to be instantiated in R.)

At this point, one point should be made absolutely clear. A moment ago I said that for there to be a surge of anger in R is just for a certain *concept* to be instantiated in R. Here I am using the word 'concept' in its *objective* sense. The word 'concept' has two quite distinct senses: a subjective or psychological sense and an objective or logical sense. Consider the sentence 'for any object x, if x falls under the concept *square* then x necessarily also falls under the concept *closed planar figure.'* This sentence says absolutely nothing about anyone's mental contents. Here the word 'concept' is being used to denote purely platonic entities, entities that exist independently of any person's mental states. This is the *objective* or *logical* sense of the word 'concept'. Now consider the sentence 'in order for a three year old to have an adequate concept of the nature of sub-atomic phenomena, he would have to be a genius.' Here the word 'concept' is being used in its *subjective* or *psychological* sense. The word 'concept' here refers to something mental, to some constituent of a human mind.

When I want to refer to concepts in the *subjective sense,* I will use the term 'concept$_s$' — note the sub-script. And I will henceforth be using the term 'concept' — no subscript — to denote concepts in the *objective* sense, i.e. to refer to a certain kind of platonic, not mental, entity. (So to refer to multiple concepts in the *subjective* sense, I will use the expression 'concepts' — once again, note the subscript. And to refer to multiple concepts in the *objective* sense, I will use the term 'concepts' — no subscript.)

Materialism holds that any mental state of affairs obtains solely *in virtue* of the fact that some physical state of affairs obtains: a set of atomic interactions. So if I feel a pain at a certain time, that happens *entirely* in virtue of the fact that, at that the same time, certain atomic interactions occurred: there is nothing to my pain over and above those interactions — just as there is no-thing to the statue over and above the clay which composes it. In general, whatever mental states of affairs there are, the nature of these states is strictly determined by the nature of the atomic interactions in the world; just as the properties of a statue at time t are strictly determined by the properties possesse[d] at time t by its constituent clay. As we noted a moment ago, for such and such

a physical state of affairs to obtain in R is just for such and such a physical concept to be instantiated in R; and for thus and such a mental state of affairs to obtain in R is just for thus and such a mental concept to be instantiated in R. So if materialism is right, then whenever a mental concept is instantiated in a certain space-time region, that is *solely because* some physical concept was instantiated in that region. This is just another way of saying that, if such and such physical concepts are instantiated in a certain region, that *necessitates* that thus and such mental concepts are instantiated in that region.

In this paper, the terms 'necessitates' and 'necessary' are meant in the strictest sense. When I say that such and such is 'necessary', I do not mean that it is causally necessary, but rather that it is *metaphysically* necessary: there is no possible circumstance in which such and such is not the case. And when I use the term 'possible' (and cognate terms: 'can', 'is able', and so on) I am not referring to causal, but to metaphysical, possibility: so such and such possible if there is some hypothetical circumstance in which such and such holds.

So a materialist must hold that, if such and such physical concepts are instantiated in R, this literally *necessitates* that thus and such mental concepts are instantiated in R — just as the truth of x *is a square* necessitates that of *x is four-sided.*

In what follows, I will talk a great deal about 'necessary relations' between concepts. What I have in mind are truths like this: *for any x, if x is a Euclidean triangle, then the interior angles of x add up to 180°.* This proposition delineates *a necessary* relation holding among certain concepts: the concepts *Euclidean triangle, interior angles,* and so on. For a relation to be necessary is for it be such that it couldn't fail to hold. There is no 'possible world' where Euclidean triangles don't have interior angles adding up to *180°.*

(Not all truths about concepts are necessary. It is probably true that *for all x, if x is a resident of Antarctica, then x cannot write a fugue.* This is a truth about the concepts *resident of Antarctica, able to write a fugue,* and so forth. But it is not a necessary truth: it is perfectly possible that tomorrow a competent fugue-writer should move to Antarctica.)

Here, in outline, is how the rest of my argument against materialism will go. We've seen that, if the mental is identical with the physical, then if certain physical concepts are instantiated in a space-time region R, this literally necessitates that certain mental concepts will be instantiated in R. Given this, suppose that [r]elations of necessitation among concepts are in fact knowable a priori. In other words, suppose that, for any two concepts (in the objective sense) C and C', if one grasps C and C' then one has all the information one needs to figure out what necessary relations hold between those two concepts. If this supposition

were in fact true, then if one knew exactly what physical facts obtained in R, one could literally *deduce* what mental states of affairs obtain in R.

I submit that this supposition is in fact true; i.e. I submit that, for any two concepts C and C', if one grasps those two concepts, one ipso facto has all the information one needs to figure out what, if any, necessary relations hold between them. I will argue at length for this admittedly controversial point.

Now it is fairly clear that, if one knows exactly what physical states of affairs obtain in region R, one cannot on that basis alone *deduce* what mental states of affairs obtain in R. (One can oftentimes *induce* this. But one can never *deduce* it. Any exceptions to this thesis prove to be merely apparent, as we will see.) From this, it follows that the mental is not literally identical with the physical. In what remains of this section I will elucidate and develop this argument.

First of all, I must prove this: If one grasps two concepts (in the objective sense of the word 'concept') C and C', then one can in principle figure out a priori what, if any, necessary relations hold between them.[10] In other words, if one grasps two concepts C and C' one ipso facto has all the information one needs to figure out what necessary relations hold between C and C'. This assertion is, of course, the essence of my argument against materialism. For expository reasons, I'll put my *full* argument for this point in the last section of this paper. But the basic idea behind that argument can be stated briefly.

Concepts are platonic entities; they are not constituents of the spatio-temporal world. Concepts *must* be platonic entities, because a concept is essentially something of which there can be *instances* (there are *instances* of the concept of triangularity, of the concept of justice, and so forth); and it makes no sense to say of some spatio-temporal entity that there are *instances* of it. It makes no sense to say that there are *instances of* Socrates or Plato.

Since they are platonic entities, concepts don't stand in spatio-temporal or *a fortiori* causal relations to one another, or to anything else. So the only thing which distinguishes one concept from another is its *constitution* — its *essential properties*. Therefore the only way that one can *identify* a concept is by its constitution — by its essential or defining characteristics. (By contrast, spatio-temporal individuals and kinds can be — and usually are — identified, not by their essential or defining properties, but by their spatio-temporal relations to one's self. This is why one can identify a certain liquid as, say, water without knowing that water is H_2O. I will elucidate this in a moment.)

10 When I say that one could 'in principle' figure this out, I mean that if one were intelligent enough, had enough time, and so on. I am abstracting from what Russell called purely `medical' limitations on the individual.

Now for any two concepts C and C', what (if any) necessary relations hold between them is determined *entirely* by the structures, the constitutions, *of* those two concepts; it is not determined by anything else; in particular, it is not determined by the constitution *of* this or that possible world. (By definition, necessary relations hold in *all* possible worlds. So they are not contingent on what goes on in this, or in any other, world.) So given that one can grasp a concept only by grasping its essential properties, it follows that, if one grasps two concepts C and C', one has all the information one needs to figure out what necessary relations hold between those two concepts.

One of the points just made should be clarified. The way we identify *spatio-temporal* objects and kinds differs (or at least *can* differ) from the way we identify platonic entities. Spatio-temporal objects obviously stand in spatio-temporal relations — in particular, in *causal* relations — to things besides themselves. So one spatio-temporal object is distinguished from the next, not only by its constitution, but also by its spatio-temporal relations. This means that it is possible, in principle, to *identify* a spatio-temporal object *without* having knowledge of its constitutional properties. This is why a three year old is able to have thoughts about water — is able to have a $concept_s$ of water — without having the faintest idea that water is H_2O and, therefore, without having the faintest idea what are the essential or defining characteristics of water. Very briefly, a three year old identifies a specimen as being water by verifying that it has a certain *causal* relation to him, not by verifying that it has a certain chemical composition: more accurately, he makes this identification by verifying that it *is of* the *same* kind as something — some specimen — to which he stands in a certain causal relation. (These obscure points will be elucidated in section IV.) But platonic entities — in particular, concepts — do not stand in spatio-temporal or causal relations to anything. So one cannot identify a concept — cannot have a $concept_s$ *of* that concept — without *(if* only implicitly or inarticulately) knowing its essential or defining properties.

Many will make the following objection to the thesis in question:

> Some propositions are both necessarily true (true in all possible world) but a posteriori (such that, to know their truth value, it is not enough to understand them: empirical work is required). Examples are: `water is $H_2O$; `light is a stream of photons'; `Hesperus is Phosphorous'. Each of these propositions is equivalent to a proposition about concepts. `Hesperus is Phosphorous' is equivalent to `the concept of Hesperushood is necessarily coextensive with the concept of Phosphoroushood'. The proposition `water is H_2O' is equivalent to the proposition `the concept *water* is necessarily coextensive with the concept *H_20.*' These latter propositions affirm necessary relations between concepts. Given that Hesperus is identical with Phosphorous, it follows that the concept of Hesperushood is necessarily coextensive with the concept of Phosphoroushood. But this relation is obviously not knowable a priori; and neither

> is the relation expressed by `the concept *water* is necessarily coextensive with the concept H_20.'

I will give a complete response to this point later on (in section IV). Right now, all I will say is this: the objection is based on a failure to distinguish between concepts in the *subjective* sense (concepts) and concepts in the *objective* sense. One can indeed have two concepts$_s$ of the same thing (or of two things that have some necessary relation to each other than identity") without being able to figure this out a priori. But whenever this happens, it is because the objects of those two concepts$_s$ are *spatio-temporal entities or kinds:* the objects of those two concepts$_s$ are never *concepts.* A concept$_s$ of Hesperus is not a concept$_s$ of a *concept;* it is a concept$_s$ of a spatio-temporal *thing:* a hunk of rock orbiting the sun — *not* a concept (in any sense of the word). The object of one's concept$_s$ of Hesperus is floating in outer space. No concept is floating in outer space. (Of course, the same is true *mutatis mutandis* of one's concept$_s$ of Phosphorous.) So when one learns that Hesperus is Phosphorous, one is not learning anything about two concepts (one is learning a lot about one's concepts$_s$: but nothing about concepts); in particular, one is not learning that two concepts are coextensive. So what one is learning is in not correctly represented by the sentence `the concept of Hesperushood is coextensive with the concept of Phosphoroushood ' That sentence, I will argue, is either nonsense or it is merely a misleading way of saying that Hesperus is Phosphorous. I will give a fuller version of this argument later on (in section IV).

We will now proceed on the assumption that, if one grasps two concepts C and C', one has all the information one needs to figure out what necessary relations hold between them. (This assumption will be discharged in section IV.) Given that this last assumption is correct, it follows that if one knows what physical concepts are instantiated in space-time region R, one can literally *deduce* what mental concepts are instantiated in R. So if one knows exactly what atomic interactions are occurring in R, this provides one with a completely adequate *deductive* basis for figuring out what mental states of affairs

obtain in R. But clearly a knowledge of what physical states of affairs obtain

11 This point may require clarification. One can have two concepts$_s$ c and c' with the following three properties: (i) the objects of c and c' are both spatio-temporal individuals or kinds; (ii) these objects stands in some *necessary* relation to each other *besides* identity; and (iii) one cannot figure out that these objects are thus related without doing empirical work. Whales are necessarily mammals. There is no possible world where something is a whale but is not a mammal. Now, a person can have a concept$_s$ of the natural kind *whale* and a concepts of the natural kind mammal and yet think that whales are fish. Such a person would not be able to learn what whales were mammals *except* by doing empirical work. So here we have a case where one has two concept; such that these objects of these concept; are (i) spatio-temporal kinds and (ii) stand in some necessary relation to one another *besides* identity and (iii) one cannot figure this out a priori.

in R does not, by itself, provide one with an adequate *deductive* basis for inferring what mental sates of affairs obtain in R.

The difference between deduction and induction must be emphasized here. If materialism is right, then if I know what physical concepts are instantiated in R, I do not just have good *inductive* evidence for what kind of mental concepts are instantiated in R; I actually have such information as enables me to *deduce* what mental concepts are instantiated in R. (I am using the term 'deduce' in the same sense that it has in the sentence `if one knows that x is a square, one can deduce that x has four sides.')

So now it is clear why materialism is false. The materialist must obviously hold that mental entities are identical with, or constituted by, certain brain-states or neural events occurring in the brain. Given this, suppose I know exactly what physical concepts are instantiated in R — i.e. that I know exactly what atomic interactions are occurring in R — where R is the space-time region occupied by someone's brain. On the basis of that knowledge, it couldn't conceivably be *deduced* what mental concepts were instantiated in R. In fact, that knowledge wouldn't even provide a decent *inductive* basis for inferring what mental concepts were instantiated in R. It surely wouldn't provide any deductive basis for such an inference. But if materialism were right, then if I knew what physical concepts were instantiated in R, I could literally *compute* what mental concepts were instantiated in R — i.e. what mental states of affairs obtain in R. So reductive materialism is false.

At this point, I should address a couple of possible objections to what I've said:

> The macro-physical facts are fully constituted by the micro-physical facts. This is incontrovertible. Now, scientists had to spend years figuring out `bridge principles' by which, given a knowledge of the macrophysical facts, one could deduce what the micro-physical facts are. In other words, scientists had to spend years discover the necessary relations that hold between these two strata of facts. You seem to be saying that it can all be done a priori — that there is some kind of *entailment* relation (an epistemically transparent necessity: *a logical* relation) between the micro-facts and the macro-facts. Obviously you are wrong.

The entailment goes from the micro-facts to the macro-facts, not *vice versa:* it you know what microphysical concepts are instantiated in R, then in principle you could on that basis alone figure out what macrophysical concepts are instantiated in R. To make an equivalent point: if you know what microfacts obtain in R you can figure out what macrofacts obtain in R. If you know the atomic facts, you can, on the basis alone, figure out what the molecular facts are; on *that* basis, you can figure out slightly higher level chemical

facts; and so on. Basically, if you know the atomic facts, you can bootstrap your way to knowledge of physical facts of the highest level — to knowledge of biological, the ecological, the geological, the astronomical (just as, if you know the location and size of each of the bricks composing a given building, you can, wholly on that basis, deduce the overall structure of the building). But the reverse is not true. This is because any given macrofact can be realized by an essentially infinite number of different types of microfacts. The event of a heart's beating in a certain way can be constituted by infinitely many different sets of atomic interactions. So there is no a priori route from the macro to the micro.

But there is an a priori route from the micro to the macro. (Of course, human knowledge begins with the macro. So we can't take advantage of this route.) This is because the micro-facts strictly and unilaterally determine the macro-facts. Although a given kind of heartbeat can be realized by infinitely many different kinds of micro-interactions, a given set of micro-interactions will allow for, at most, one kind of heartbeat.[12]

Another objection:

> 'It is often said that mind is an "emergent property" of matter. By this it is meant that mental activity is (i) physical but (ii) represents an "irreducible novelty" in the physical world. So if this view is correct, then the mental is physical even though it is not constituted by atomic interactions. Of course, if mental states of affairs aren't constituted by atomic states of affairs, then there is no reason why, if such and such atomic states of affairs obtain, that should entail that thus and such mental facts obtain.'

This position is incoherent. If the mental facts constitute an `irreducible novelty' with respect to the physical facts, this means that what the mental facts are is not strictly determined by what the physical facts are. This, in turn, means that mental entities and phenomena are not identical with, or constituted by, the physical entities and phenomena. (If x is fully constituted by y, then the x-facts cannot be `irreducibly novel' with respect to the y-facts.) This, in its turn, means that mental entities just *aren't* physical. So the idea that mind is an `emergent property' — i.e. is irreducibly novel with respect to the physical — while itself being physical is self-contradictory. Actually, this position is very close to the position that I have called *non-reductive* materialism.

12 See Nagel [1] p.352.

Property Dualism

Our main goal is to explain the systematic correspondence that obtains between the mental and the physical. We have seen the problems $_{\text{inherent}}$ in both Cartesian dualism and in reductive materialism. Aware of these problems, some philosophers have proposed a kind of compromise between these two doctrines: mental properties are distinct from physical properties, but all mental properties belong to physical objects. So the pain I feel is *a property* of my brain, but it is an irreducibly mental property. This view is known as 'property-dualism'. [13]

Property dualism explains the correspondence between the mental and the physical by saying: mental properties co-occur with physical properties because both types of properties belong to physical objects. Pain always succeeds such and such physical stimuli because those stimuli cause thus and such neural events, and pain is a property (albeit an irreducibly mental property) of those neural events.

There is an obvious problem for property dualism. It is widely agreed these days that objects just *are* concomitances of properties. A given object — e.g. a particular rock — is not something in which properties *inhere;* it is not something *underlying* the various properties which it possesses. Rather, it is the sum of its properties. (I am using the word `sum' loosely, of course.) So to say that two properties — e.g. hardness and roundness — `belong to the same thing' just *is* to say that they co-occur: it is *not* to say that they are `glued' or `affixed' to the same substrate. Property dualism says that the mental and the physical co-occur because they `belong to the same thing'. But to belong to the same thing — to `inhere' in the same object — just *is* to co-occur. So property dualism in effect just says: *mental and physical properties co-occur because they co-occur* — and this, of course, is utterly trivial.

To make all of this clear: Property dualism is supposed to explain *why* the occurrence of mental properties attends that of physical properties and *vice versa.* And its answer is: the occurrence of mental properties attends the occurrence of physical properties, and *vice versa,* because (in some cases) mental properties and physical properties `belong to the same things'. But, as we've just seen, for two properties to belong to the same thing just *is* for them to co-occur — i.e. just *is* for the occurrence of the one to attend the occurrence of the other. So property dualism reduces to the vacuous statement that *the occurrence of mental properties attends that of physical properties, and vice*

13 Colin McGinn is a property dualist. See McGinn [1].

versa, because the occurrence of mental properties attends that of physical properties. This statement, of course, has no explanatory content. So property dualism is null and void as a solution to the mind-body problem.

IV. A Positive Solution

We have seen some reason to hold that the mental and the physical do not interact and that mental activity is not physical activity. But it also patently obvious that there is some intimate and genuine connection between mind and matter, between mental and physical activity. The apparent responsiveness of the mental to the physical, and *vice versa*, is not coincidence. How are we to account for this correspondence?

So far as I can tell, there is only way left to do this. What we call `physical objects' are projections or representations of objects whose essential properties are mental. *I am not advocating any kind of idealism.* Rather, I am saying (i) that, existing in complete independence of human minds, there exist objects that are just like rocks, chairs, and trees, except that these objects are composed of *mental* entities (mental entities that do not belong to any human mind); and (ii) that what we call `physical objects' are projections or isomorphs of these other objects (roughly: physical objects are to these other objects what Kantian *phenomena* are to *noumena).*

If I am to defend this hypothesis, I must be permitted a brief epistemological digression. We obviously learn about the physical world through our sense-perceptions. Now there is some reason to believe that perception, and perception-based theories, apprise us only of the structure of physical objects, and not of their non-structural ('constitutive') properties. To begin with, sense-perception makes us aware of two kinds of properties: structural and phenomenal. When you see a chair or a rock or a tree, you see something that (a) has a certain `bulk, figure, and [state of] motion' (to use Locke's expression). But this is not *all* you see: for those `primary' — those purely structural — proper-ties are necessarily `clothed' in so-called `secondary' properties color, odour, firmness, taste, coolness etc. Indeed, quite clearly, an object that had *no* so-called secondary properties would not be perceptible at all. So physical objects, as *they are given to us in sense-perception* (leaving aside for the moment how they might be in themselves), might be thought of as structural skeletons whose flesh is phenomenal properties.

So *prima facie* sense-perception seems to apprise us of non-structural properties, viz. phenomenal properties. But, depending on how we think of them, phenomenal properties are either purely subjective — *sensations* experienced

projectively as properties of objects — or as purely structural features of objects. (So the property of being red is a micro-structural property: the property of having a certain micro-configuration or — what may be closely connected, both causally and conceptually — of reflecting light of certain wavelengths.) If the phenomenal properties of objects are just sensations, then `phenomenal' properties are not properties of objects *at all;* so that, in seeing an object as having a certain phenomenal property, we are not aware of any property that it *really* has — and sense-perception is not revealing any genuine non-structural properties of objects. On the other hand, if e.g. the property of being red or being sweet is just a micro-structural property, then — it follows trivially — in seeing an object as having this or that phenomenal property — as experiencing it as red or sweet — one *is* learning something about the object; but one is, after all, learning (ultimately) that it has a certain structural property. So either

way, physical objects as given to us in sense-perception, are purely structural entities.

There is another way to establish this same point. Your perceiving an object as being red (or sweet or pungent...) obviously involves your having some kind of *sensation;* some kind of subjective, sensual response to these objects. (To use the current philosophical jargon, your perceiving something as sweet or red...necessarily has a `phenomenology': there is, in Nagel's phrase, `something it is like' to perceive an object.) Let us focus for a moment on those sensations of ours that are involved in the experience of objects' so-called secondary properties. Trivially, either there is, or there is not, a consistent relationship between those sensations and the properties of objects that set them off. If there is *no* consistent relationship, then in having those sensations, we are learning *nothing* about physical objects. On the other hand, if there *is* a consistent relationship, then what we are learning about are purely structural properties.

An example is no doubt needed to elucidate this: consider the sensation you have when you experience the sweetness of an object. Let S be that kind of experience. For the sake of argument, suppose that the various objects in response to which you experienced S had *nothing* in common *other* than their disposition to make you have S: more specifically, suppose that those objects had *no* structural properties in common — that their respective micro-structures varied without limit. In that case, if you had S in response to tasting two different types of cake, you could not, on that basis, legitimately infer *anything* about the properties of those two food-items (that is, you could not make any inferences *other* than the purely trivial one that they both caused you to have S). On the other hand, if those two items *do* have anything in common (other

than having a disposition to cause people to experience S), that will inevitably be some micro-structural (or micro-causal) property. Either way, what is learned from an object on the basis of its phenomenal properties is, *if anything at all,* some purely structural feature of that object.

The same holds for any other secondary property. Consider the kind of sensation you have (under a given set of conditions) in response to `red' objects — to stop signs, tomatoes, fire-engines etc. Let R be this kind of sensation. Now suppose that, given any two `red' objects, those two objects had *no* micro-structural or micro-causal properties in common. In that case, in experiencing R in response to an object, you would obviously be learning *nothing* about it (except some completely anthropocentric fact about it, viz. that it makes you have a certain kind of sensation — that it makes you *feel* a certain way, essentially: which obviously doesn't qualify as knowledge *about the object* in any proper sense). On the other hand, if your having R in response to two different objects *does* correlate with some objective property of those objects, that property will inevitably be some micro-structural/micro-causa[1] property. So what one is learning about an object in experiencing it as `red' is, *if anything,* that that object has some *structural* property.

Basically, phenomenal properties — if they are properties of objects at all, rather than just the way we feel in response to objects — are exactly like the property of heat. For something to be 78° is (very roughly) for the molecules that make it up to have a certain mean kinetic energy. Obviously, in response to objects having that temperature, we *could* in principle have any sensation at all. (So consider the way that e.g. a body of water that is 78° makes one feel; let S I be that kind of sensation. And consider the way a body of water that is 48° makes us feel; let S2 be *that* kind of sensation. Obviously, through *e.g.* neuro-surgery, one could be *made* to feel S2 in response to water that is 78° and SI in response to water that is 48°. The possibility of that reversal clearly doesn't imply that temperature is something subjective.[14]) Although the perception of something as being hot is closely bound up with the having of a certain kind of sensation, it is clear that, in so far as one is learning anything about objects on the basis of that kind of sensation, one is learning that objects have some kind of structural property in common. If the objects which caused that kind of sensation varied without limit in respect of their micro-structural properties — if, what is equivalent, our having that sensation did not correlate at all consistently with purely structural properties of the objects which caused it — then that sensation would ipso facto be non-epistemic: we would learn

1 4 See Kripke [1], p.129.

nothing about the (extra-mental) world through our having it. The same is true of our experience of red. The having of sensations of type R (the kind typically experienced in response to apples, blood, fir engines etc.) either (i) *does* correlate with certain microstructural properties of the objects that cause R; or (ii) it *does not* so correlate. If (ii), then in having R we are learning nothing about those objects on the basis of having R. If (i) then what we are learning about those objects on the basis of having R is that those objects have certain *structural* properties.

To sum up, we've seen two reasons to believe that sense-perception apprises us only of *structural* properties of objects. In sense-perception things are given to us as having two kinds of properties — structural properties — 'bulk, figure....' — and phenomenal properties. But phenomenal properties are either (a) purely subjective or (b) *if non-subjective* then purely structural — are basically `bulk, figure, and motion' in disguise.

What about physical objects as known to us through hypothesis based on sense-perception? It is a commonplace in the philosophy of science that theoretical entities are known *exclusively* in terms of the structural properties. Our knowledge of what e.g. quarks are is *at least* as formal (structural), as our perceptual knowledge of chairs. I say `*at least as* abstract' and not `*more* abstract' since, as I have just said, our perceptual knowledge of chairs is already completely formal. As for why exactly *why* theoretical entities are known only as regard as their `formal' properties — this is a delicate question. Presumably the answer has to do with the large role played by *analogy* in the postulation of theoretical entities. To put the matter extremely roughly, theoretical entities — while more basic *ontologically* than directly perceived entities — are less basic *epistemically;* and we seem to grope our way towards a grasp of the micro-foundations of our world by positing tentative analogies with what is directly perceived. Analogy is, of course, a form-preserving — not a content-preserving — operation. (For two things to be analogous is specifically for them to have a common *form,* not common non-formal properties.). So given that theoretical entities are known to us analogically, they can only be known as regards their formal/structural properties.

It was considerations like these that Russell to conclude that physical objects are known only `as regards their space-time *structure.'* (Russell's emphasis.) Elsewhere, again basing himself on reasoning at least somewhat similar to that just set forth, Russell writes:

> `[W]e have found it necessary to emphasize the extremely abstract character of physical knowledge, and that fact that physics leaves open all kinds of possibilities as to the intrinsic character of the world to which its equations apply. There is nothing in

> physics to prove that the physical world is radically different in character from the mental world...The only legitimate attitude about the physical world seems to be one of complete agnosticism as regards all but its mathematical properties.' [15]

In any case, with regard to the idea that our knowledge of e.g. electrons etc. might be *less* formal than our knowledge of chairs and rocks — that idea is, plainly, a non-starter. So given that our knowledge of directly perceived objects is purely structural, so *a fortiori* is our knowledge of hypothetical (in particular, microscopic and sub-microscopic) objects. To sum up, there is some reason to believe that our knowledge of the perceived world is purely structural (phenomenal properties either become absorbed into the structural part of the world or they drop out of the physical world altogether) and that so a *fortiori* is our knowledge of the microstructural basis of it.

But external objects must have non-structural, or (as I will henceforth say) `constitutive', properties. There cannot, ultimately, be disembodied structures. (Disembodied structures exist: mathematics studies them. But such things have no causal powers, and are therefore not among the constituents of the spatio-temporal world.) This is because structures consist in relations *between* entities (a simple object has *no* structure). Not all entities can consist in relations that hold between other, simpler entities. Such a conception implies a vicious regress. As Wittgenstein put it, a world all of whose constituents have structure is a world that has no substance, i.e. that contains no objects whatsoever.[16] So we must hold that, in addition to having phenomenal properties and structural properties, a physical object also has *constitutive* properties.

Could we conceivably discover, through natural scientific investigation, the constitutive properties of physical objects? If what we said about sense-perception is correct, then we could not. We know the physical world only in so far as it has a certain structural similarity to the phenomenal world. So, at most, we know its structure. We don't know what has this structure. Basing himself on reasoning similar to that just set forth, Russell once wrote:

> I conclude that, while mental events and their qualities can be known without infer- ence, physical events are known only as regards their space-time structure. The qua- lities that compose such events are unknown - so completely unknown that we cannot say either that they are, or that they are not, different from the qualities that we know as belonging to mental events."

15 Russell [2] pp. 270-271.
16 *Tractatus Logico-philosophicus.*
17 Russell [1] p. 247. Elsewhere Russell writes:

> [W] e have found it necessary to emphasize the extremely abstract character of physical knowledge, and that fact that physics leaves open all kinds of possibilities as to the intrinsic character of the world to which its equations apply. There is nothing in physics to prove that the physical world is radically different in character from the mental world...The only legitimate attitude about the physical world seems to be one of complete agnosticism as regards all but its mathematical properties.' Russell [2] pp. 270-271.

Suppose that the constitutive properties of physical objects are *mental* in nature. If that were the case, then we wouldn't have to explain how mental entities came into existence. For mental entities would simply have existed *ab initio.* We don't currently feel that we must explain how physical objects came into existence; we take it for granted that physical objects are the ultimate constituents of the spatio-temporal world, in both the causal and the mereological senses of `ultimate'. If, in fact, mental entities had this status, then we wouldn't have to explain them. (The existence of mental activity would be beyond explanation, just as the existence of physical activity is currently reckoned to be beyond explanation.)

In a moment I will actually *reject* this hypothesis. But first I must state its merits: for the view that I will endorse can only be understood *in terms* of this hypothesis. Indeed, the former might — if only in a loose, technically inaccurate way — be seen as but a variant of the latter.

The hypothesis in question (that the `constitutive' properties of some physical entities — presumably brains— are mental in nature) explains the concordance that subsists between the mental and the physical; i.e. it explains the apparent responsiveness of the one to other. It is fairly clear, as a matter of empirical fact, that every event in a person's mental life is accompanied by some change in the state of his brain. It is also clear that every change in a person's brain, above a certain order of magnitude, is accompanied by some change in his mental life. Whenever such-and-such happens in my brain, I feel pain; and whenever I feel pain, such-and-such happens in my brain. We can't explain this by saying that such-and-such brain-events produce thus-and-such mental events or *vice versa* (for brain-events cause, and are caused by, physical events alone: the existence of mental activity cannot damage the causal integrity of the physical world); or by saying that mental events are physical events. So how are we to explain why (e.g.) pain is always accompanied by such-and-such physical events?

If the *constitutive* properties of those physical events consisted in pain, then it would be perfectly understandable why this correspondence obtained. Basically if physical phenomena have as their *constitutive* properties those mental phenomena that always accompany them, then it is no wonder that those mental phenomena and those physical phenomena are always conjoin-

ed. It is not to the discredit of this theory that, no matter how thoroughly We examine physical nature, we never discover mental entities to be among its constituents; for, as we have established, we cannot possibly know, through an examination of physical nature, what its constitutive properties are.

I hear an objection to this theory:

> You cannot coherently countenance this theory. You spent a great deal of time trying to prove that the mental and the physical are not identical and that neither constitutes the other. Therefore you cannot now say that the mental constitutes the physical.

The interlocutor is right. In its current form, I cannot countenance theory just set forth. But I can countenance a slightly rectified version of it. The rectification I am proposing will not be ad hoc; it will follow from independently arrived at truths concerning the concept of physicality.

What do we mean by the term `physical object'? What do we mean when we characterize something as `physical'? One answer is this: an object is physical if it falls within the scope of one of the so-called `physical sciences' — physics, chemistry, and biology — and the sub-disciplines that they comprise. (This definition appears circular: for it defines `physical object' in terms of `physical science'. But in a moment we will make it non-circular.) It seems that if something is such that it couldn't *conceivably* be the object of study of one of these disciplines — and, therefore, that it couldn't be discovered by one of these disciplines — then surely it wouldn't be physical. It also seems that, conversely, if something does (at least conceivably) fall within the scope of these sciences, then it is physical. Even materialists hold this. A materialist will indeed hold that physical objects are studied by a discipline other than biology, chemistry, and physics: for he holds that pains, tickles, beliefs, and so on are physical and are studied by psychology, which is distinct from physics, chemistry, and biology. But the materialist holds that pains, tickles, beliefs and so on are identical with things studied physics, chemistry, and biology — that they are identical with brain-states and brain-structures. The materialist is willing to concede that *if* so-called mental entities (pains, beliefs, etc.) were not identical with the things studied by physics, chemistry, and biology, then indeed they wouldn't be physical. So the materialist holds that mental entities are identical with physical entities only because they are identical with the kinds of things studied by the so-called physical sciences. (Of course, the non-reductive materialist holds that mental entities are physical and yet are not identical with the kinds of things studied by physics, chemistry, or biology. But we have seen that non-reductive materialism is not a form of materialism at all; it is Cartesian dualism. So, from now on, by `materialism', I will mean

only *reductive* materialism.) So for something to be physical is for it to be the kind of thing that could, at least potentially, fall within the scope of physical sciences.

But this definition of `physical' is circular, *unless* we can find some way to define the term `physical sciences' *without* employing the term `physical' (or any synonym). In other words, if we define a `physical' object as one that is studied by the `physical sciences', and we then define the `physical sciences' as those sciences that study `physical' objects, then our definition of `physical' is circular, and therefore worthless. But if we define the term 'physical object' to mean the kind of thing studied by the `physical sciences', and we then go on to define the latter term *independently* of the term `physical', or any synonym thereof, then our definition will be acceptable. This is what I now propose to do.

Under what circumstances does something fall within the scope of the so-called physical sciences? There are two possible circumstances. (i) If an object is *sense-perceived* it falls with the scope of the physical sciences. Trees and rocks are studied by the physical sciences because they are sense-perceived. (ii) An object that is not sense-perceived (e.g. an atom) will fall within the scope of the physical sciences so long as the *empirical basis* for knowledge of it lies exclusively in sense-perception. Atoms, quarks, and force fields are not sense-perceived. But they are studied by the physical sciences because the empirical basis of our knowledge of them lies exclusively in sense-perception.

Condition (i) is straightforward. But condition (ii) requires elucidation: What does it mean to say that the `empirical basis' of our knowledge of a thing lies `exclusively' in sense-perception? Every substantive belief about the spatio-temporal world has to have some basis in either sense-perception or in what is sometimes called `introspection'. Of course, not every belief about the spatio-temporal world (and possibly not any of them — though I think this might be an overstatement) follow *directly* from sense-perception. We believe that atoms exist. But this belief doesn't follow directly from sense-perception; we don't really see atoms. We infer that they exist *on the basis* of what we see. This inference consists in our bringing to bear certain canons of logic (broadly defined) to directly perceived data. This inference — like all inferences to matters of spatio-temporal fact — thus has both a purely rational basis and an empirical basis. The empirical basis, of course, lies in certain sense-perceptions.

Some beliefs about the spatio-temporal world have an empirical basis that lies, at least partly, in something other than sense-perception. If I believe that I am in pain, or that I am sad, or that I believe that snow is white, this belief

will *not* result from sense-perception, or even (in most cases) from inferences made on the basis of sense-perception. It will have an empirical basis, at least a partial one, in some non-perceptual modality. Let us refer to this other modality, whatsoever its nature might be, as `introspection'.

Many of our beliefs about the spatio-temporal world have an empirical basis *both* in sense-perception *and* in introspection. If I see Joe writhing and groaning, I will conclude that he is in pain. My belief is obviously based partly on sense-perception (my sense-perceptions of Joe's body). But it isn't *wholly* based on sense-perception. Unless I had actually *had* pain — unless I knew about pain in some way other than through sense-perception — I wouldn't have any idea what pain was; I wouldn't have the concept of pain; and I therefore couldn't infer, from a knowledge of Joe's physical state, that he was in pain. This seems to be true, not just of pain, but of all mental entities — even of mental states, like desire, which have strong conceptual ties to certain kinds of behaviour. Unless I had actually had emotions, beliefs, intentions, desires, and so on, I wouldn't really know of such things; and I therefore couldn't impute them to others. So with regard to our knowledge of other people's minds, and of the unconscious contents of our own minds, the empirical basis of this knowledge lies partly in sense-perception and also in introspection. With regard to knowledge of our own minds, the empirical basis this knowledge usually, though probably not always, has its basis solely in introspection.

The *physical* sciences are those whose empirical basis does *not* lie in introspection; they are those sciences whose empirical basis lies exclusively in sense-perception. Now at last we have a non-circular definition of what it is to be physical: something is physical if it falls within the scope one of physical sciences, and therefore could in principle be *discovered* by one of those sciences; and a science is a physical science if its empirical basis lies in sense-perception, and not to any degree in introspection.

This definition is neutral between materialism and dualism. The materialist holds that emotions, sensations, perceptions, and so on, are physical precisely because they are identical with things that fall within the scope of physics, chemistry, and biology: things identical with, or constituted by, displacements of atoms, brains-states, neural events, and so on. The materialist is perfectly willing to admit that *if (e.g.)* pains do not fall within the scope of one of these sciences — that *if* pains are not identical with (say) neural events — then indeed pains are not physical. So this definition is compatible with material-ism. But this definition is also compatible with dualism; it allows for the possibility that some things cannot be learned of through the physical sciences. So in defining the concept of physicality in this way, we haven't pre-

judged the truth of any doctrine concerning the mind-body problem.

Given this definition of what it is to be a physical object, we can fix up our earlier faulty solution to the mind-body problem. For something to be physical is for it to fall within the scope of one of the physical sciences. Of course, something falls within the scope of the so-called physical sciences only if it is in principle *discoverable* by one of those sciences. Now, as we noted, the physical sciences apprise us only of *structure.* Therefore it follows that anything non-structural is non-physical: the concept *physical* object is a structural concept. At the same time, we noted that, for purely conceptual reasons, there cannot be disembodied structures. So certain non-structural, certain *constitutive,* properties are required to `flesh out' the structures that physical objects *are.* But these constitutive properties are not themselves physical. For the physical is that the empirical basis for knowledge of which lies wholly in sense-perception; and any knowledge that is exclusively perception-based is knowledge of structure. So we must say that physical objects are in some way *associated* with certain constitutive properties, but that physical objects don't actually *have* these properties.

So physical objects are `associated' with certain constitutive properties while not actually *having* them? But how exactly is this association to be conceived? We must conceive of it, I think, as follows. The neural events that accompany pain are representations or projections of mental events. To put this in Kantian terminology: mental entities are the *noumena,* and physical entities are the *phenomena.* Physical objects are how mental objects are given to us *in sense-perception* and through theories whose empirical basis lies entirely in sense-perception Physical objects are representations of mental objects. The constitutive properties of the physical world are mental. What we call `physical objects' are analogues of these properties.

This is the exact opposite of what is usually held: physical objects are usually held to be basic; mental objects are held to be derivative, either causally (interactionism) or ontologically (supervenience), of matter. But we have also seen that there is simply no way to extract mind from matter.

However, if we take mind as basic (not *our* minds, but mentation in general), there doesn't seem to be any impossibility in principle in explaining the existence of physical objects. The chair will never produce any mental state in me; it will never, in particular, produce any perception in me. Photons will bounce off of the chair; some of these will disturb certain bodily surfaces of mine (my retinas). These disturbances, in their turn, will produce certain disturbances of my optical nerves. These in their turn will precipitate certain neural events, which in their turn may produce all manner of other physical events. But nowhere in

this concatenation of physical events is there room for anything mental: no matter how assiduously we study all these physical processes, our examination will refer us only to more physical processes. We don't need to hypothesize the existence of mental events to explain this processes. In fact, we couldn't possibly find any way of inserting them into these processes. (Mental characteristics cannot coherently be attributed to physical entities.) My perception of the chair is commonly held to be causally dependent on the chair. But the chair seems incapable of creating anything other than physical events.

Now imagine the following scenario. There is some object in outer space, some object that exists independently of my mind and everyone else's. This object (we might call this the `noumenal' chair) has the exact same *structural* features as the physical chair, i.e. *of* the chair *qua* thing knowable through sense-perception and intellectual extensions thereof (we might call this the `phenomenal' chair). But the noumenal chair is composed *of* mental entities. These mental entities affect other mental entities, whose phenomenal counter-parts are certain physical particles. These entities precipitate other mental events, whose phenomenal counterparts are certain disturbances of certain bodily surfaces of mine. These mental events, at last, precipitate a perception *of* the chair: the phenomenal counterpart of this perception is some brain state or series *of* brain states.

Physical states and interactions mirror mental states and interactions. We have the physical interacting with the physical and, running alongside, the mental interacting with the mental. The parallelism is explained by saying that the physical is a kind *of* projections of the mental. We have already seen why the mental cannot be a representation or projection of the physical.

My desire to move my arm *doesn't* move my arm. Rather, it moves the `noumena' corresponding to my arm. The movement of my arm is a phenomenal projection of that movement. My pain always accompanies certain kinds of neural events *not* because my pain is identical with such disturbances; nor because my pain causes, or is caused by, such disturbances; but because such disturbances are the `phenomenal form' of my pain.

Of course, the argument just set forth isn't valid unless our analysis of what it is to be physical was valid. We said, basically, that something is physical just in case the empirical foundation for knowledge of it lies in sense-perception. Now, some people would object to this, arguing as follows:

> For something to be physical, it is enough that (i) it is in space and (ii) it has causal powers. (Condition (ii) is needed to rule out things like the equator and space-time points — ideal, and therefore non-physical, entities which are in space.) For something to be physical, it isn't necessary that it satisfy any other conditions — e.g. that it be discoverable through physics, chemistry, or biology.

It is pretty easy to show that our concept of physicality is richer than this objection makes it out to be; that something could be in space and have causal powers, and yet fail to be physical. Really, we already saw why this is so when we discussed non-reductive materialism.

Suppose that our mental states are (i) in space and (ii) they have causal powers over each other, but (iii) they don't have causal powers over *paradigmatically* physical objects — over the kinds of objects that fall within the scope of physics, chemistry, and biology (things like atoms, molecules, kidneys, and so forth). Under these circumstances, would mental entities qualify as physical? Suppose we said they did; and suppose that, in keeping with this, we used the term `physical' to refer both to paradigmatically physical entities *and* to things like pains, tickles, and perceptions. I submit that, if we did this, we would thereby render the term `physical' ambiguous. Given that this term is not currently ambiguous, it follows that the current meaning *of* the term `physical' doesn't apply to entities *merely* in virtue *of* their being in space-time and having causal powers.

A point made by Hilary Putnam may be *of* service here.[18] Suppose that, here on Earth, there was some substance that were *not* composed *of* H_2O — whose microstructure was, in fact ,quite different from H_2O — but whose surface properties were like those of H_2O, and that therefore was, from a purely pragmatic perspective, equivalent to water. Suppose that the microstructural differences between this substance and H_2O were not discovered until 1980. Under these circumstances, we would almost certainly refer to H_2O and to this other substance with the same word. (Suppose that this word was `water') Putnam asks: under these circumstances, would the word `water' be ambiguous? His answer is: yes.

Putnam is right. If the term `water' denoted substances that had different microstructures, and that therefore didn't have the same *law-like connections* to other physical entities, this word would simply be ambiguous. For our language to do justice to the structure *of* the world, it would have to come up with two different words for these two different substances.

Suppose that, after 1980, language did so, and that the two words were `water1' and '$water_2$'. The words `water1' and '$water_2$' would not denote two different *species* of a single genus. In other words, they wouldn't denote two different varieties of the same substance. They would denote altogether different substances. Generality must be distinguished from ambiguity. The term `red' covers various different colours; it covers maroon, burgundy, candy-

18 See Putnam [1].

apple red, fire-engine red, and so on. But the term `red' isn't ambiguous; for maroon, burgundy, etc. are different versions of the same property. The microproperties in virtue of which an object is light-red are similar to those in virtue of which an object is burgundy. These microstructures, in their turn, determine to a large extent the behaviour of the object in question. So, in virtue of being two different shades of red, two objects will have a great deal in common *other than their being red.* Their redness will correlate with other similarities between them; it will correlate with their having other properties in common, where these other properties are the kinds in terms of which scientific, law-like explanations are made. So it isn't a short-coming of language that it refers to maroon, burgundy, etc. with one word; that language does so is actually to its credit: for the use of a single word to cover these different properties embodies an *insight:* the insight that these properties are related.

But, in the above thought-experiment, the term `water' (before 1980) would refer to altogether different things; it doesn't refer to different varieties of a single kind of thing. **In** virtue of having different microstructures, the two substances in question would behave very differently in different contexts; and these differences would not be *systematic.* To make a related point, al-though $water_1$ and $water_2$ have the same phenomenal properties, this commonality wouldn't correlate with other commonalities; it wouldn't correspond to their having *properties in common apart from the aforementioned phenomenal properties. So* their having these phenomenal properties in common would be explanatorily sterile; it wouldn't correspond to law-like, systematic connections between the two substances. The property of being water, would be explanatorily *disjoint* from the property of being $water_2$. By contrast, given two objects, each of a different shade of red, it would in a wide variety of physical contexts be possible to trace the differing behaviours of those two objects to the different degrees to which possessed a certain property, where this property was what was responsible for their being shades of red.

Let us bring these reflections to bear on the topic at hand. If the term `physical' covered *both* the paradigmatically physical *as well* as things that couldn't interact with the paradigmatically physical, then that word would be like the word `water' in the thought-experiment; it wouldn't be like the term `red'. Again, suppose mental entities existed in space-time and had causal powers (over each other), but not over paradigmatically physical entities. If we referred to such objects as physical, then the `physical' would no longer constitute an explanatorily unified domain; it would cover two disjoint domains. The differences between these domains would not be *systematic;* they would

not be attributable to the different degrees in which they possessed some single characteristic. The term `paradigmatically physical' carves nature at the joints; it picks out a unified, systematically interconnected class of entities. If the term `physical' covered both the paradigmatically as well as things that, while being in space and having causal powers, couldn't affect then the paradigmatically physical, then the term `physical' wouldn't pick out a unified, systematically interconnected class of entities; it wouldn't be a natural-kind word. It would therefore be ambiguous, in the way that, in the above thought-experiment, `water' was ambiguous. But the term `physical' isn't ambiguous; it is a natural kind term, albeit an extremely general one. Therefore the term `physical' doesn't cover entities that can't affect the paradigmatically physical. So for something to be physical, it is not enough that it have causal powers and be in space; it must also be *paradigmatically* physical. So our definition of physicality is vindicated.

Of course, in response to this, one might say:

> How do we know that the term `physical' picks out a unified domain? Don't we have to wait for science to be completed before this thesis can be fully verified? For all we know, in ten years we'll discover some massive breach in the causal structure of the so-called "physical" world, in which case it would turn out this word was ambiguous — like the term "water" in the above thought-experiment.

This objection is correct. But it has no real bearing on what we've said. Suppose it turns out the word `physical' is ambiguous; in other words, suppose that the so-called `paradigmatically physical' world turned out *not* to be a causally unified domain. Given this, if we were to countenance the application of this word to objects that *were* not covered by the term `paradigmatically physical', this would *add* an ambiguity to the term `physical' that it didn't *already* have. So even though this term was already ambiguous, allowing it to refer to objects other than the paradigmatically physical would make it even *more* ambiguous; it would therefore cease to have its current meaning. This means that its current meaning covers only the paradigmatically physical.

Let us sum up what we've said so far. The mental and the physical *seem* to be responsive to each. Cartesian dualism can't explain why this is. Materialism *could* explain it. But we have seen that materialism is false. The solution to our puzzle is to be found through careful scrutiny of the *concept* of physicality. The physical is that which is to be known on the basis of sense-perception, and not on the basis of introspection. Sense-perception, and the theories built thereupon, apprise us only of structure, not of content. But there cannot be disembodied structures; there must always be content. If we assume that the

contents *corresponding to*, but not identical with, physical objects are mental, then we have a solution to the mind-body problem.

Before proceeding we should consider an important objection to the argument just set forth:

> You say that for something to be physical is for it to be such that the empirical basis for knowledge of it lies exclusively in sense-perception. So you are defining 'physicality' in terms of `sense-perception.' But there seems to be no way to define the concept of sense-perception except in terms of the concept of physicality. So your definition of physicality is circular.
> Why must `sense-perception' be defined in terms of `physicality'? For me to perceive the chair, it is necessary that the chair *physically* affect me in certain ways; it is necessary that my mental state be the result of disturbances of certain sensory surfaces of mine that were precipitated, ultimately, by the chair. (If the chair has no causal affect on me at all, then no matter what the subjective character of my mental state — no matter what kind of mental image I am having, for example — I will not be having a perception of the chair by virtue of being in that mental state.)

Indeed, perception is an inherently *causal* notion. But not all causation is *physical* causation; there is mental causation as well. (Operating in conjunction with each other, your state of thirst and your perception of the ice water cause you to have an intention to reach out and grab the glass of ice-water. This is a case where two mental entities interact to produce a new mental entity: a case of mental causation.) And the kind of causation involved in sense-perception needn't be —and, I submit, isn't — *physical* causation. The objector is right to say that, for me to perceive the chair, I must have some causal relation the chair. But the objector has misdescribed the nature of that relation. That relation is, I submit, to be thought of as follows. The chair qua physical object — i.e. the chair qua thing with such and such *structural* properties — does not affect my mind in any way. But the chair qua object with such and such *constitutive* properties *does* affect my mind. (Of course, the same point applies to all the entities and processes mediating between the chair and my mind. Qua physical objects — qua things possessed of such and such structural properties — these intervening entities and processes do not affect my mind. But qua objects possessed of such and such constitutive properties, these intervening entities and processes and entities *do* affect my mind.) We saw reason to believe that the constitutive properties of physical objects are mental in nature. If this is correct, then the chair's effect on me is a case of purely *mental* causation — a case of one mental entity's affecting another. So in this we can reconcile the fact that, for me to perceive the chair it is necessary that the chair effect me with the fact that nothing mental can produce or affect anything physical.

IV. Dualism and Conceivability

Part 1: Conceivability and Possibility

The argument against `reductive materialism', given above, goes through only if it is the case that, for any concepts C and C', if one grasps those two concepts, one ipso facto has the all information one needs to figure out what necessary relations hold between them. I gave a brief argument for this thesis; now I'd like to flesh out that argument. (I must do so because this thesis is *highly* controversial.) The best way to begin my defence of this thesis is to consider an objection to it (here I am quoting a passage given earlier):

> Some propositions are both necessarily true (true in all possible world) but a posteriori (such that, to know their truth value, it is not enough to understand them: empirical work is required). Examples are: `water is $H_20$; `light is a stream of photons'; `Hesperus is Phosphorous'. Each of these propositions is equivalent to a proposition about concepts. `Hesperus is Phosphorous' is equivalent to `the concept of Hesperushood is coextensive with the concept of Phosphoroushood'. And the Proposition `light is a stream of photons' is equivalent to the proposition `the concept water is coextensive with the concept $H_20$. `These latter sentences express necessary relations between concepts — they express `necessary relations', in your terminology. Given that Hesperus is identical with Phosphorous, it follows that the concept of Hesperushood is necessarily coextensive with the concept of Phosphoroushood'. But this relation is obviously not knowable a priori; and neither is the relation expressed by *the concept* water *is necessarily coextensive with the concept* H_20.

It cannot be denied that some necessarily true propositions are a posteriori. But such propositions are *not* about concepts. Necessarily true propositions are either a priori or they are not about concepts.

To see why this is so, we must make it clear how it is that there can be a posteriori necessary propositions in the first place. Hilary Putnam's classic thought experiment (slightly revised) will help us do this.[19] Let Twin-Earth be some planet that is qualitatively just like Earth — a planet whose past, present, and future consist of events and states of affairs just like those composing Earth's past, present, and future — *except* that on Twin-Earth the substance in oceans, bathtubs, and so on, is *not* composed of H_2O, but has some other chemical composition. (Let xyz be this chemical.) xyz is phenomenally just like water (H_20), and it serves the same practical functions as water. There are important microstructural differences between H_2O and xyz, but these do not become apparent except under narrowly defined experimental conditions. Of course, even though water and xyz are superficially very similar, xyz is not

19 See Putnam [1]

water: after all, water is H_2O, and xyz is not H_2O. Given all of this, suppose that Joe is a cognitively normal three year old living on Earth; and suppose that Twin-Joe is Joe's counterpart on Twin-Earth. In terms of their *internal* or *subjective* characteristics, Twin-Joe and Joe are qualitatively identical. (In other words, if you consider only those properties of theirs that can be defined or individuated independently of objects external to them, Joe and Twin-Joe are exactly alike.) But Joe has thoughts about *water* — about H_2O — and he never has thoughts about xyz; and Twin-Joe has thoughts about xyz, and never about water. When Joe says `water is transparent' he is expressing a thought about what is in fact $H_2O$; whereas when Twin-Joe says `water is transparent', *he is* not expressing a thought about H_2O, but about xyz. Why is it that, even though Joe and Twin-Joe are qualitatively identical so far as their *internal* properties are concerned, Joe has a concept$_s$ of H_2O and *not* of xyz, whereas Twin-Joe has a concept$_s$ of xyz and not of H_2O?

Surely the answer is this: Joe is *causally* connected in a certain way to H_2O but *not* to xyz; whereas Twin-Joe is causally connected in a certain way to xyz but not H_2O. So Joe's concept$_s$ of water — that which enables him to single out water in his mind, to have thoughts about water — is constituted, in part, by some *causal nexus* mediating between himself and H_2O, or some specimen thereof. In general, one's concepts$_s$ of spatio-temporal entities are oftentimes (arguably always) constituted by *causal relations* mediating between oneself and the entity in question.

Given this last point, it is clear how it is that one can have two concepts$_s$ that apply to the same object without being able to figure this out a priori, i.e. without being able to figure this out on the basis of what is `in one' s head'. In such a case, in order for one to figure out that these two concepts$_s$ had the same object, one would, in effect, have to find out what lay at the other end of two separate causal chains; in such a case, finding out that two concepts$_s$ had the same object would be tantamount to finding out that two causal sequences terminated in the same object; and this, plainly, is not something that can be done a priori. Part of Joe's concept$_s$ of water — his means of cognitively locking onto water — in effect is a certain stretch of extra-cranial spatio-temporal reality. Of course, such a stretch is not transparent to Joe — is not such that its depths can be plumbed through thought alone — in the way that a concept$_s$ lying entirely within Joe's head would be. Consequently, Joe could have two concepts$_s$ of (e.g.) water — or of Venus or of Tully- without being able to figure this out *a priori:* for, in effect, these concepts would consist, in part, in stretches of the extra-cranial spatio-temporal world, and of course the properties such a stretch cannot be excogitated *a priori.*

Now, one's concept$_s$ of a concept in the *objective sense* cannot possibly be constituted, to any degree, by one's causal relation to that concept (to denote concepts *in the objective sense*, I will simply use the word `concept': no subscript); for concepts are not spatio-temporal, and therefore don't stand in spatio-temporal or (a fortiori) causal relations. Concepts (in the objective sense) are not among the constituents of this or that possible world. (It would be more correct to say that they exist *between* worlds than to say that they exist *in* worlds.) Since they are not spatio-temporal, one does not enter into causal relations with them; *a fortiori* no concept$_s$ that one has of a concept involves a stretch of the spatio-temporal world. (There are, I fully grant, apparent counter-examples to this. But these counter-examples are *merely* apparent, as I will try to show.) To sum up, one cannot identify — cannot pin down in thought — a concept by its spatio-temporal relations, since a concept has no such relations.

So how is one to identify, to pin down in thought, a concept? Two concepts differ from each other only in respect of their *constitutions. So* one can distinguish one concept from the next only by its constitution. So one must *grasp* the constitution — the essential or defining properties — of a concept to have a concept$_s$ of it. As we noted earlier, given any two concepts C and C', what necessary relations hold between them is determined entirely by their constitutions. So if one grasps these two concepts, one has all the information one needs to figure out what necessary relations hold between them.

Now we can respond directly to what the interlocutor said. Any given a posteriori sentence *seems* to be equivalent to some sentence about concepts. For example, `heat is molecular motion' *appears* to be equivalent to the sentence `the concept *heat* is coextensive with the concept *molecular motion.'* This appearance is an illusion. First of all, as we've noted, to have a concept$_s$ of heat is not to have a concept$_s$ of a concept; in particular, it is not to have a concept$_s$ of a concept that applies, in any possible world, to heat. Now a concept of *heat* is just such a concept: it is a concept that applies in any possible world to all and only instance of heat. How does one get from having a concept$_s$ of heat (the phenomenon in the world) to having a concept$_s$ of the concept *heat* (that platonic entity which, in any possible world, applies to all and only instances of heat)? Having a concept$_s$ of heat means only that, *in this world*, one can identify instances of heat. As we've seen, one often identifies spatio-temporal individuals and kinds by their spatio-temporal relations; in particular, by their causal relations to one's self. Grasping a concept of heat — in other words, having a concept$_s$ *not* of the spatio-temporal phenomenon of heat, but of a concept of heat — means being able to pick out heat in *hypothetical* worlds. Now one cannot identify a phenomenon in a hypothetical world

as heat by verifying that it stands in some causal relation to one's self; for phenomena in hypothetical worlds have no such relations to one. If one cannot identify a phenomenon by its causal relation to one's self, then one must identify it by its *constitution.* Therefore one can identify instance of heat in *hypothetical* worlds only by knowing what the *constitution* of heat is. So in order for one to grasp a concept of heat, one must know what the constitution of heat is: one must know that heat is molecular motion (if, in fact, that is what it is). So, in fact, one cannot grasp the proposition `the concept heat is coextensive with the concept *molecular* motion' without recognizing it as true. So this proposition is necessary *a priori,* not necessary a posteriori. So the proposition `heat is molecular motion' does not correspond to any necessary *a posteriori* proposition about concepts: It corresponds only to some necessary *a priori* proposition about concepts. Of course, what we've said about the sentence `heat is molecular motion' is true of all a posteriori necessary sentences. Although any given necessary a posteriori sentence corresponds to some proposition about concepts, the latter will always be a priori. So the existence of necessary a posteriori truths in no way counterexamples our thesis that, if one grasps two concepts C and C', then one ipso facto has all the information one needs to figure out what necessary relations hold between them.

There is one important objection to this thesis:

> An analysis of a concept gives the essential or defining properties of that concept. An example of an analysis is: a circle is a closed planar figure of uniform curvature. Analyses are informative. This shows that one can grasp concepts without grasping their essential or defining characteristics.

There are two possible reasons why analyses might be informative. One is that they tell us things that we *simply didn't know.* The other is that they make explicit knowledge which was previously implicit; or, at any rate, that they in some way transform existing knowledge. There are a couple of good reasons to take the second of these two views.

Suppose that Joe grasps the concept *circle,* but he doesn't know (explicitly) that a circle is a closed planar figure of uniform curvature. In principle[20] ,

20 By 'in principle' I mean 'assuming Joe were intelligent enough, had enough energy' and so on. Joe himself may not have the intelligence to arrive at a correct analysis of the concept *circle* on the basis of what is 'in his head'. But what is preventing Joe from being able to arrive at such an analysis is not a lack of empirical information. It is a lack of intelligence. Given any one who grasps the concept. *circle,* if that person is unable to arrive at a correct analysis of that concept, it is *not* because of a lack of empirical information. In this essay I make this point by saying that 'in principle' anyone who grasps that concept could arrive at a correct analysis of it without doing empirical work: so the 'in principle' here means (roughly) 'all other things being equal'.

Joe obviously doesn't have to do empirical work to arrive at a correct analysis of this concept — to arrive at the knowledge that a circle is a closed planar figure of uniform curvature. (It is fairly clear that, in principle, no one who grasps this concept need do empirical work to arrive at a correct analysis of it.) This means that Joe has enough information already — has enough information `in his head' — to arrive at this analysis. Let us refer to Joe's knowledge of this information as *inf. So,* in virtue of having *inf,* Joe knows of some proposition (or set of propositions) P that logically implies the proposition that a circle is a closed planar figure of uniform curvature.

One point about *inf* must be made explicit: Joe's possession of inf is what enables Joe to think about the concept *circle. So* Joe's $concept_s$ of the concept *circle is identical* with his possession of inf. Why this is so becomes clear when we lay out the relevant facts. Joe's $concept_s$ of the concept *circle* is what enables Joe to think about the latter. (This is just a truism.) Joe doesn't (in principle) have to do empirical work to arrive at an analysis of the concept *circle.* He need only reflect on what is `in his head', so to speak. Naturally, to arrive at such an analysis he must reflect on his own $concept_s$ of the concept *circle;* and that is the *only* thing he must reflect on to arrive at this analysis. By definition *inf* is Joe's knowledge of such propositions as imply a proposition giving the analysis in question. So to arrive at the analysis in question, Joe must reflect on *inf;* and there is nothing besides *inf* that he must reflect on. It follows that *inf is* identical with Joe's $concept_s$ of the concept *circle.* It will become clear in a moment why this seemingly trivial point is important.

As we noted, we must assume that Joe knows of some propositions that *imply* a proposition (or set of propositions) giving an analysis of the concept *circle.* (If we didn't make this assumption, it would be inexplicable how it is that Joe is able to arrive at a correct analysis of this concept without doing empirical work.) Given that Joe has a $concept_s$ of the concept *circle,* can we coherently assume that Joe has knowledge only of such propositions as *imply* an analysis of the concept *circle* but that he doesn't (at some level) have knowledge of this analysis itself? It doesn't seem so: this becomes clear as soon as we reflect on the difference between knowing a proposition P and merely knowing some proposition that *implies* P.

An example may be helpful. The solution to the continuum problem is given by some sequence of propositions. So this solution is a kind of platonic entity. (We might even think of it as a concept.) Now, to solve the continuum problem — to figure out what the aforementioned sequence of propositions was — I wouldn't (in principle) have to do empirical work: there is enough information `in my head' for me to do this. (In fact, empirical information

would be totally irrelevant to any effort to solve this problem: it is a problem of mathematics, not of empirical science.) But I don't know what the solution to the continuum problem is: I am not *acquainted* with this solution. I have no *direct* knowledge of this solution. (I have, at most, what might be called 'knowledge by description' or `indirect knowledge' : I know some of the conditions that a platonic entity would have to satisfy to qualify as a solution, but I don't know *which* platonic entity does so.) So given only that one is in possession of such information as enables one to figure out what a certain platonic entity is, it doesn't follow that one is acquainted with that entity. Now, Joe is quite plainly *acquainted* with the concept *circle;* he has a kind of *direct* knowledge of it. (Joe is just as capable of having thoughts that are *about this concept* as is the best of mathematicians; so he is no less acquainted with this concept than the mathematician. The difference is that the mathematician knows *more* about this concept than Joe.) In any case, he grasps the concept *circle* with a directness and an immediacy that sharply distinguishes it from my grasp (if such it can be called) of the solution to the continuum problem. So it cannot be that in virtue of having *inf,* Joe *only* has knowledge of such propositions as *imply* that proposition that a circle is a closed planar figure of uniform curvature. For if that were the case, then Joe's grasp of the concept *circle* would be as indirect, as mediated, as my grasp of the solution to the continuum problem. It must be that, in virtue of having *inf,* Joe actually grasps the truth that a circle is a closed planar figure of uniform curvature. So when Joe learns that a circle is a closed planar figure of uniform curvature, what is happening is hitherto implicit or inarticulate knowledge of Joe's is being transformed made explicit and articulate. In general, analyses are informative *not* because they provide knowledge where previously there was ignorance *tout court,* but because they make explicit knowledge that was previously implicit.

This argument is, I think, borne out by pre-theoretical intuitions. Consider a paradigm case of ignorance. A month ago, someone stole my tennis racket. I simply don't know where it is. (It could be in some other country right now.) Can someone grasp a concept and be ignorant of its essential properties the way I am ignorant of the location of my tennis racket? Intuitively there seems to be a difference. It would seem that oftentimes (if not always) when someone is given an analysis a concept that he grasps, he *recognizes* in that analysis what he knew all along. If this is correct, it would suggest that analysis transfigures existing knowledge — that it makes explicit knowledge that was previously implicit. To sum up, both intuition and argument indicate that analysis makes explicit knowledge that was hitherto implicit. Consequently the fact that analyses are informative in no way casts doubt on my contention

that, for one to grasp a concept, one must grasp its essential or defining properties.

Let us finish up this section by considering one last objection:

> As you admit, oftentimes one's concept$_s$ of something spatio-temporal involves a causal connection to that thing. But what is the nature of that causal connection? Presumably it is this: the thing in question causes you to have certain mental states. But if the thing in question causes you to have certain mental states, then the physical does cause the mental, contrary to what you've tried to show here.

We've already seen how to deal with this sort of objection.[21] It is the chair's *noumenal* or *constitutive* properties that affect my mind — that cause me to have certain mental contents. These properties are, we agreed, purely mental. It is not the chair *qua* physical thing — not the chair *qua* thing possessed of such and such *structural* properties — that affects my mind. Again, it is the chair *qua* thing with certain constitutive — certain non-structural, certain mental — properties that affects my mind. So the causal chain mediating between myself and the chair — the causal chain constituting (in part) my concept$_s$ of the chair — is a purely *mental* chain. To be sure, corresponding to this mental chain is a phenomenal chain — a chain consisting of the physical or structural properties associated with the aforementioned mental or constitutive properties. But this phenomenal chain is not *per se* what constitutes my epistemic rapport with the chair; it is just a concomitant of that rapport, a phenomenal projection of it.

Inevitably some will make the following objection to the argument just given:

> "Your argument goes through only if there is no such thing as *the* concept of water or *the* concept of Socrates. But surely this is mistaken. Consider the following propositions:
> (i) *if x falls under the concept* Socrates, *then x falls under the concept* human.
> (ii) *if x fall under the concept* water *then x does not fall under the concept* is an element.
> Surely (i) and (ii) are true; and they are true in virtue of facts about the concepts *Socrates* and *water."*

My first response is this: the objector is putting much too much stock in the fact that natural language permits certain expressions to be substituted for others. The rules of English syntax do permit the substitution of `x, and only x, falls under the concept *Socrates* and x is bald' for `Socrates is bald'. But from

21 See the end of section III.

this fact, surely, no conclusions can be drawn about ontology; surely we cannot read metaphysics off of grammar. Surely the convertibility of `Socrates is human' with (i) tells us only about grammar — about the rules governing syntactical permutations — and nothing about the fundamental features of reality. In particular, it doesn't show us that there is such a thing as *the* concept *Socrates.* And, I submit, there is no such thing.

To begin with, if there *were* such a thing as *the* concept *Socrates,* that concept would be `object-dependent' (or 'object-involving'), i.e. it would have a spatio-temporal individual for its content. (I will use the terms `object-involving' and `object-dependent' interchangeably.) But the very idea of an object-dependent concept —a concept that has, for example, Socrates himself or water itself as a constituent — is an absurdity.

But before we can see this, we must make it as clear as possible just what an object-dependent concept is supposed to be, and why such concepts are thought to exist. Consider the proposition:

(*)Socrates drank hemlock

Socrates himself is an actual constituent of M. The idea will become more clear if we consider a slightly different proposition:

(**) *there was a philosopher of antiquity who exceeded all others in philosophical ability and any philosopher answering that description drank hemlock.*

(**) is *made true* by the fact that Socrates was the greatest philosophers of antiquity and that he died of drinking hemlock. But (**) does not have Socrates himself as a constituent. That very proposition does not depend for its truth on *Socrates'* having such and such characters. That very proposition would have been true if Socrates had never existed, and *some other person* was the greatest philosopher and died of hemlock poisoning. So Socrates himself is not a constituent of (**).

By contrast, (*) would *not* be true if *anyone other than Socrates* had the property of being the greatest philosopher of antiquity and dying of hemlock poisoning. (*) depends for its truth on Socrates specifically having those properties. So Socrates himself figures in the truth-conditions of (*) and, in as much as propositions are internally or essentially related to their truth-conditions, Socrates himself can be said to be a constituent of that proposition.

Now it seems reasonable to say that propositions are built entirely out of concepts (though I deny this below); and in the case of (*) these concepts would presumably be *Socrates, hemlock,* and so forth. Given this last point, and given — what we saw a moment ago — that Socrates himself is a constituent of (*), it very much seems to follow that the concept *Socrates* has Socrates himself for its content. Presumably Socrates manages to be a constituent of (*) only by way of his involvement in the concept *Socrates. So* the concept *Socrates* has an actual constituent of the spatio-temporal world for its content, this constituent being Socrates. So the concept *Socrates is object-involving,* as it is generally put (it is object-involving *with respect to Socrates).* (If a concept has only a platonic object for its content, it does not count as `object-involving', even though platonic objects are objects of sorts.)

Object-involving concepts, it is alleged, can have natural kinds for their contents — their contents needn't always be spatio-temporal *individuals.* This is supposed to follow by an analogue of the argument just given. Consider the proposition

(***) *water freezes at 32°*

This proposition is object-involving with respect to the natural kind *water.* What does this mean? The best way to see what this means is to contrast it with a proposition that is not object-involving with respect to water:

(****) there is some substance that human beings bathe in that freezes at 32°.

(****) is *made true* by the fact that we bathe in water and that water freezes at 32°. But (****) doesn't depend for its truth on *water's* having a certain freezing point. If there were some other substance with a freezing point of 32° that we bathed in, then (****) would be true. But (***) is not like this: for (***) to be true, it is necessary that *water* — specifically *water* — be such that we bathe in it and that it have a certain freezing point. So *water itself* figures in the truth-conditions of (***). And , in as much as propositions are internally or essentially related to their truth-conditions, water itself — the natural kind — can be said to be a constituent of (***).

It seems reasonable, if not truistic, to say that (***) is built out of various *concepts,* these being *water, 32°,* and so forth. So given this last point, and given that water itself — the natural kind — is a constituent of (***), it seems to follow that the concept *water* has the natural kind *water* for its content. Presumably that natural kind succeeds in being a constituent of (***) only by way of its involvement with the concept *water. So* that concept must have the natural kind *water* for its content. Thus, the concept *water is object-involving;* for it has a natural kind for its content.

The concept *triangle* is not object-involving; for its content is some purely platonic object, not an individual, and not some natural kind/scattered object like water. The same is true of various other concepts: *number, justice, truth, knowledge, implication.* This completes our exposition of the reason why object-dependent (object-involving) concepts were held to exist.

We will now see that, although object-dependent *propositions* exist, there

is no such thing as an object-dependent *concept.*

We've seen that *if* there were such a thing as *the* concept *Socrates,* that concept would have Socrates himself for its content: that concept would, in effect, *be* identical with the individual Socrates. (The same is true *mutatis mutandis* of the concept *water:* if there were such a thing, it would be identical with the natural kind.) But a concept is not a part of the spatio-temporal world. A concept is a mode of presentation of a property. The concepts *closed figure of uniform curvature* and *closed shape whose peripheral points are equidistant from a given point* pick out the same property — that of being a circle — even though they are different concepts. (The two concepts have the same referent — the property of circularity — but different senses. Better, they *are* different senses.) But an *individual* — e.g. Socrates — is not a mode of presentation. To say otherwise would be sheer nonsense. The natural kind *water is* not a mode of presentation. So there is no such thing as *the* concept of Socrates or *the* concept of water: for spatio-temporal individuals and kinds are not modes of presentations of properties and are therefore not concepts.

There is another way of refuting the objector. Concepts are ultimately things of which there are *instances.* There *are instances* of the concept *round.* There are no *instances* of the concept *Socrates.* It is meaningless to say that Socrates himself, the individual, is instantiated by something. (This corresponds to the Aristotelian point that Socrates cannot be *predicated* of anything, whereas baldness can.) Now if were such a thing as *the* concept of Socrates, that thing would, as we have seen, be identical with Socrates himself. So the idea that there is such a thing as *the* concept of Socrates is committed to the nonsensical view that there can be instances of Socrates himself — the nonsensical point that Socrates can be *predicated* of things. (Admittedly, some concepts cannot have instances, e.g. *round-square.* But any such concept is built up out of concepts that can have instances — in this case, *round* and *square. So ultimately* concepts are things of which there can be instances.)

Of course, there are concepts$_s$ (note the subscript) *of* Socrates. In other words, there are mental contents that have Socrates for their objects. But there is not such a thing as *the* concept *Socrates.* Socrates is an individual, not a concept.

Undeniably, the proposition *Socrates is bald is* object-dependent with respect to Socrates. For this proposition to be true, it is necessary that Socrates, and no one else, be bald: so Socrates himself is implicated in — is a part of — that proposition. But I deny that *Socrates is bald* has *a concept* of Socrates for a constituent. It has *the individual,* not a concept thereof, for a constituent. The error in the argument given above, for the existence of object-dependent concepts, lay in the assumption that *Socrates is bald* is constructed entirely out of

concepts. It is not: it is constructed out of concepts (e.g. *bald) and* an individual (Socrates). Once it is seen that Socrates, but not a concept thereof, figures in *Socrates is bald,* then there is no reason to countenance the idea of *the* concept *Socrates — the* same argument *mutatis mutandis* showing that there is no such thing as *the* concept *water* or *the* concept *Plato,* and so on. So there is no legitimate transition from `Hesperus is Phosphorous' to `the concept of Hesperushood is necessarily coextensive with the concept of Phosphoroushood' . In general, object-involving propositions cannot be transformed into propositions about object-involving concepts. There are no such concepts; the only concepts that exist are *not* object-involving. So the objector's point fails; and if a statement is necessary and a posteriori — e.g. *water is* H_2O — it is made true, not by the structure of *concepts,* but by the structure of *spatiotemporal* entities. Finally, if a concept is necessary and it is made true by the structure of concepts — e.g. *the interior angles of a Euclidean triangle add up to 180°*— it is a priori. So necessary relations among concepts *can* always be excogitated a priori; the existence of necessary a posteriori truths does not bear against this.

Part 2: Another argument for dualism

Nonetheless *even if it is granted* that there is such a thing as *the* concept of Socrates (or, what is more or less the same, *identical with Socrates)* and *the* concept of heat, *the* concept of water, and so forth — even if this is granted, an argument for dualism can easily be constructed. In what follows, I will, in deference to the object, operate on the assumption that there is such a thing as *the* concept of Socrates, *the* concept of heat, and so on.

The old argument for dualism was basically this. What is not counter-conceptual — i.e. what is not ruled out by the structure of concepts — is possible. Triangles can be green because *x is a triangle* is logically consistent with x *is green.* Now *xis (e.g.) a pain or a belief that 2+2=4* does not logically entail *x is a brain event. So* it is logically possible that beliefs, pains, etc. should be distinct from brain-events (or any other kind of physical event).

The next step in the argument is this. Identity holds *necessarily.* If A *can* be distinct from B, then *A must* be distinct from B. Proof: B obviously doesn't have the property that it can be distinct from B. So if A has the property that it can be distinct from B, then A has a property that B does not have and so, by Leibniz's law, A is simply *not* B. (There are some apparent counterexamples to this principle — e.g. (*) `the inventor of bifocals *is* identical with the first postmaster general, but the former didn't *have* to be identical with the latter.' But Russell and Kripke showed that (*) is not an identity at all; it says merely:

some one individual x had two sets of properties —x had the property of being a postmaster before anyone else and x also had the property of being an inventor of bifocals before anyone else.)

Once it is granted that identity is necessary, and that x *is a pain (or a belief that 2+2=4...)* doesn't entail x *is a brain-event,* it follows that pains, beliefs, etc. are not brain-events: dualism proved.

These days, of course, the counter-argument is to deny that logical possibility entails actual ('metaphysical') possibility:

> "Given only that *xis a pain (or a belief that 2+2=4...)* is logically consistent with xis *not a brain event,* it does *not* follow that pains are necessarily *distinct* from brain events. Why not? Well, *x is water* is logically consistent with *x is not H_20,* but we know from chemistry that water is H_20, and couldn't be anything else. Some necessities are *a posteriori,* and logical possibility therefore proves nothing as to *actual* ('metaphysical') possibility."

The conclusion is vastly overdrawn. Surely logical possibility *sometimes* indicates actual possibility. *x is a triangle* is logically consistent with *x is green,* and this *does* mean that there could be green triangles. At the same time, x *is Hesperus* is logically consistent with *x is not Phosphorous,* but nothing that is Hesperus could not be Phosphorous. And x is *water and x is not H_2O* is logically consistent, but this does *not* mean that water could be something other than H_2O. What is the relevant difference among these cases?

Whenever there is a disparity between logical possibility and actual possibility — or between actual necessity and logical necessity — that is because the possibilities/necessities in question are underwritten by *object-dependent concepts.* And whenever a necessity is underwritten wholly by object-independent concepts, actual necessity/possibility coincides with logical necessity/possibility.

When a proposition is necessarily true, and its truth is underwritten by the structure of object-independent concepts, that proposition is a priori. (Compare *squares have four sides.)* When a proposition is necessarily true, and its truth is underwritten by the structure of object-dependent concepts, that proposition is a posteriori. The truth of *Hesperus is lovely* necessitates the truth of *Phosphorous is lovely.* But this necessity is underwritten by object-dependent concepts *(Hesperus, Phosphorous) is* therefore a posteriori — is not a *logical* necessity (is not an entailment). The same is true *mutatis mutandis* of the necessary connection between x *is water* and x *is H_20.* On the other hand, the concepts of triangularity and of two-sidedness are not object-involving, and that is why the necessary relation between x is *a triangle* and *x has more than two-sides is* epistemically transparent.

We've seen some examples that support this thesis that, where object-independent concepts are concerned, logical possibility *does* correspond to actual possibility. But is there a more general justification for this position?

There is. By definition, object independent concepts do not have any part of the spatio-temporal world for their contents. Now whether some kind of necessary relation holds between two concepts C 1 and C2 has to do entirely with the constitutions of those concepts. Necessary relations hold in all possible worlds. So whether a necessary relation holds between two concepts can-not be contingent on what happens in this or that world, and must therefore have to do entirely with the structures of those two concepts. *(x is a triangle* necessitates the truth of *x has more than two sides* because of something about the constitutions of the concepts *triangle* and *two sides.* The same being true *mutatis mutandis* in the case of *x is water* necessitates the truth of *x is H_2O* — even though the relevant facts about the constitution of *water* and H_2O are not epistemically transparent.) So given two *object-independent* concepts Cl and C2, no empirical work — no investigation of the spatio-temporal world — is needed to know in what necessary relations they stand with respect to each other. Only purely *conceptual* — purely *a priori* — work is involved.

But where object-dependent concepts are concerned, this is not the case. One has to do empirical work to know the constitutions of such concepts —for such concepts have stretches of the empirical world for their contents. So it cannot typically be known a priori in what necessary relations object-dependent concepts stands with respect to each other.

To sum up, where object-independent concepts are concerned, logical possibility coincides with actual possibility; where object-dependent concepts are concerned, this is not the case. I will use this fact as a way of arguing for dualism — as a way of reviving the conceivability argument described a moment ago. So let's say that Cl is some mental concept (i.e. some concept such that, if x falls under it, then x is ipso facto mental: the concept *belief that 2+2=4 is* such a concept). And let's say that C2 is some physical concept (i.e. some concept such that, if x falls under it, then x is ipso facto physical: *has a positive electrical charge* would be such a concept). If Cl and C2 are object-independent, and *x falls under CI* is logically compatible with *x does not fall under C2,* then things falling under Cl are *not* identical with things falling under C2. For, to reiterate, where object-independent concepts are concerned, logical possibility/necessity coincides with *actual* possibility/necessity. So if x *falls under CI* is logically consistent with *x does not fall under C2,* then it is possible for things falling under C1 to be distinct from things falling under C2. More formally, for any x, any y, if x falls under Cl and y falls under C2, it is

possible that x is distinct from y. And since identity holds *necessarily,* it follows that x *is* distinct from y.

In this section I will show, first, that at least some mental and physical states of affairs can be described entirely in terms of object-independent concepts; and, second, that the concepts involved are *logically* consistent with some mental entities not being physical entities. Since, where object-independent concepts are concerned, logical consistency implies actual possibility, it can be inferred that some mental entities really can be distinct from physical entities. And since *x can* be distinct from y only if x is distinct from y, it follows that some mental entities are not physical.

Let us begin. First of all, what is a good test of whether a concept C is object-dependent or not? Remember what we said earlier about Joe and twin-Joe. Whether Joe grasps water (or in this context: the *concept* of water) as opposed to twin-water (or in this context: the *concept* of twin-water) is *not* wholly determined by what is Joe's mental contents, narrowly individuated, are. Rather, it is determined by what Joe's contents, narrowly individuated, are *plus* what kind of environment Joe is in along with *how* he is embedded in that environment. The earmark of an object-dependent concept is this: one's grasping such a concept is not a function merely of what one's mental contents, narrowly individuated, are: it is a function also of what one's causal liaisons to the external world are.

So if possession of a concept C is *not* sensitive to facts about one's causal liaisons to the external world, then C is object-independent.

Given this, suppose the following. Joe and twin-Joe are exactly alike as far as their mental contents *narrowly individuated* are concerned. But Joe and twin-Joe are in utterly different physical environments. Now Joe, like any cognitively normal human being, grasps the concept *belief that 2 + 2 = 4*. (In other words, he knows what it is for somebody to believe that 2+2=4.) Given the facts, as we've just described them, does it make any sense to suppose that twin-Joe does *not* have the concept *belief that 2 + 2 = 4 ?*

Surely not. To illustrate this, let us consider the most extreme realization of the facts as we've just described them. Suppose the following. Joe is an ordinary human being on Earth. Twin-Joe is a brain in a vat. But twin-Joe's mental life, narrowly individuated, is just like Joe's. (In other words, twin-Joe's mental life — considered apart from his being a brain in a vat, and apart from the all the causal facts associated therewith — is just like Joe's.) Surely twin-Joe, despite his unfortunate predicament, knows just as well as Joe what it is to have the belief that 2+2=4. Surely twin-Joe can manipulate this concept (i.e. the concept *believes that 2 + 2 = 4)*, and knows its application-conditions,

as well as Joe. So we must conclude that the concept *belief that 2 + 2 = 4 is* object-*independent*.

An exactly analogous argument can be given to show that various other mental concepts — e.g. *desire for a meaningful life, ticklish sensation, love of poetry* — are object-independent. (Some may have misgivings about applying what we said about *belief that 2 + 2 = 4* to *is a pain* or *is a ticklish sensation.* These misgivings are unwarranted, and I deal with them below.)

Not *all* mental concepts are object-independent; for some mental concepts implicate object-dependent concepts. For example, the concept *belief that Socrates was smarter than Plato* implicates the object-dependent concepts *Plato* and *Socrates.* So even if two people are exactly alike as far as their mental contents, narrowly individuated, are concerned, it might be the case that one of them has the concept *belief that Socrates was smarter than Plato* while the other does not; for possession of this concept involves, not merely having certain contents, but also having certain causal liaisons to one's environment. Twin—Joe, being a brain in a vat, will not have these causal liasons, and will not grasp this concept, even though Joe does.

But plainly *some* mental concepts are object-independent. The concept *belief that 3 is greater than 2 is* object-independent: a brain in a vat could, in principle, grasp this concept as well as anyone. Suppose that twin-Gauss is mentally just like Gauss, except that twin-Gauss is a brain in a vat. Surely twin-Gauss has the same mathematical acumen as Gauss — has the same intelligence about number as Gauss. And surely twin-Gauss is just as able to apply the associated mental concepts — e.g. *believes that there are infinitely many primes* — as Gauss.

Let us now turn to physical concepts. Obviously some physical concepts are object-dependent, e.g. *identical with water, identical with heat.* But it seems to me that the most *fundamental* physical concepts cannot be object-dependent. The purpose of any science is to provide as complete a description of the objects falling in its purview as possible. The purpose of e.g. theoretical physics is to provide as accurate, as fine-grained, and as complete a description of the states of affairs falling in its scope. Now to the extent that a proposition is object-involving — i.e. to the extent that it involves object-involving concepts — it has (by definition) actual *objects,* rather than *descriptions* of those objects, for its content. The proposition *Socrates was bald* is object-involving: Socrates himself is a constituent. Inevitably, a proposition that has Socrates himself as a constituent is (ceteris paribus) less information rich than one that contains *a description* of Socrates. If you replace *Socrates,* in the just mentioned proposition, with some description of, say, the mental and physical events

associated with Socrates, the resulting proposition will be incomparably more fine-grained, more information-heavy than the former. (I am not saying — what Kripke proved false — that *Socrates was bald* is synonymous with some statement of the form *the unique such and such was bald.* All I am maintaining is that, when objects occurring in propositions are replaced with descriptions, then — holding everything else constant — the resulting proposition, though perhaps not synonymous with the original, is quite obviously much richer in information than the former.) In general, there can be no doubt that, in so far as *objects* are constituents of a proposition, rather than *descriptions* of objects, that proposition is of a lower degree of complexity, and of information-richness, than it would otherwise be.

When objects occur in propositions, they occur as simples — even though objects *per se* are not simple. The proposition *Socrates was bald* has a very simple structure — one of the form *a has phi* — even though Socrates himself was very complex. For whatever reason, Socrates' complexity is not implicated in the proposition *Socrates was bald.* And this point applies to *any* object that becomes a constituent of a proposition. Water is complex; it has a molecular and atomic structure. But this complexity is not implicated in the proposition *water freezes at 32°;* that proposition has a maximally simple form — it has the form *a has phi.* To sum up, whenever a spatio-temporal entity figures as a constituent of a proposition, its *complexity* — its internal structure — is not implicated in that proposition; spatio-temporal objects, in propositional contexts, are utterly *simple.*

But when *a description* occurs in a proposition, all of its complexity *is* implicated in that proposition. So *the greatest bald philosopher of all time was Greek is* more complicated a proposition — and therefore, if true, more information-rich — than *Socrates was bald;* the same being true for any other proposition that results from *Socrates was bald* by replacing *Socrates* with a description. In general, in so far as a proposition involves object-dependent concepts, it is not as information-rich as it could be.

Now theoretical physics is concerned with generating maximally precise — maximally information-rich — propositions about sub-atomic phenomena. An idealized theoretical physics really just is a set of maximally accurate such propositions. As we've just seen, to the extent that a proposition employs object-dependent concepts, that proposition is not as fine-grained as it would otherwise be. So the foundational concepts of theoretical physics — the concepts in terms of which the most precise and exhaustive description of sub-atomic reality are to be couched — must be *object-independent.* These must not comprise actual chunks of the spatio-temporal world; for in so far as they

do, they do not do justice to the internal structure of those chunks.

So the propositions of a physics which captures the fine-grain of the sub-atomic world must be couched in object-independent concepts. The concepts of such a physics must be *object-independent.*

Now let us move on to the next phase of the argument. As we noted earlier, for a state of affairs to occur in space-time region R is just for some concept to be instantiated in that region. For there to be a particle of such and such charge and mass, moving at such and such velocity in region R, just is for the concept *particle with such and such charge and mass moving with such and such velocity* to be instantiated in R — the same being true *mutatis mutandis* for any other state of affairs that might occur in R (or any other region). So for some microphysical state of affairs to obtain in R is just for some micro-physical concept to be instantiated in R.

Everyone agrees that whatever physical states of affairs hold in a region R supervenes on what microphysical states of affairs hold in R. The biological, the chemical, the geological, and so on, supervene on the microphysical. This is equivalent to saying: what physical concepts are instantiated in R — what biological or chemical or geological... concepts are instantiated in R — supervenes on what microphysical concepts are instantiated in R.

The materialist holds that whatever mental states of affairs are instantiated in R supervenes on (is strictly determined) by what physical concepts are instantiated in R. This is equivalent to saying: for the materialist, whatever mental concepts are instantiated in R supervenes on what microphysical (atomic and sub-atomic) concepts are instantiated in R. So the materialist holds that some *necessary* relation holds between certain microphysical and mental concepts: some relation of the form *when CI is instantiated in R, C2 is also instantiated in R,* where C1 is a microphysical concept (e.g. *object with such and such charge...)* and C2 is a mental concept (e.g. *belief that 2+2=4).*

We've observed that, where object-independent concepts are concerned, necessary relations among those concepts can be excogitated a priori; since (by the definition of object-independent), such concepts do not have any component of the spatio-temporal world for any of their content, and are therefore to be known through non-empirical — purely conceptual or a priori — labour.

We've also noted that many mental concepts — e.g. *belief that 2+2=4* — are object-independent. *And* we've noted that the foundational concepts of theoretical physics are, or ought to be (ultimately), object-independent: so any maximally *precise* statement of a sub-atomic state of affairs will be one that uses only object-independent concepts — will have the form `C is instantiated in R', where C is an object-independent microphysical concept.

So if materialism is right — if any belief that 2+2=4 *is* identical with some physical state of affairs — then the following must be true: if the concept *belief that 2+2=4 is* instantiated in R, that is in virtue of the fact that some object-independent microphysical concept C (or, more likely, set of such concepts) is instantiated in R. In other words, the instantiating of C in *R necessitates* the instantiating of *belief that 2+2=4* in R. (More plainly, if C is instantiated in R, that necessitates that there be a belief that 2+2=4 in R.)

Further, because C and *belief that 2+2=4* are object-independent, the following holds: the fact that C is instantiated in R would *entail* — would *logically*, not (just) metaphysically, necessitate — that *belief that 2+2=4* was also instantiated in R. For recall that, if materialism is right, the truth of *C is instantiated in R* necessitates the truth of *the concept* belief that 2+2=4 is *instantiated in R.* And because the concepts C and *belief that 2+2=4* are object-independent, this necessity is an *entailment;* it is *a logical* necessity — one that can be excogitated a priori.

But I find it very hard to believe that there is any purely logical entailment from *C is instantiated in R,* where C is an object-independent microphysical concept, to *there is a belief that 2+2=4 in R.* It follows that a belief that 2+2=4. It is very hard to believe that the instantiating of some microphysical concept *logically necessitates* the instantiating of the concept *belief that 2+2=4.*

Let us now put all the pieces together and close the argument. Recall that where object-independent concepts are concerned, necessary relations are the same thing as entailment relations. There is, presumably, no entailment from *C is instantiated in R* to *the concept* belief that 2+2=4 being instantiated *in R.* (In other words, there is no entailment from *C is instantiated in R* to *there is a belief that 2+2=4 in R.)* Both C and *belief that 2+2=4* are object-independent concepts. So the just mentioned lack of entailment coincides with the absence of a necessary connection. So there is no *necessary* relationship between C's being instantiated in R, on the one hand, and there being a belief that 2+2=4 in R — where, once again, C is any microphysical concept. This, in turn, means that a belief that 2+2=4 is not *identical with* or *supervenient upon* the occurrence of any microphysical state of affairs in R (for identity and supervenience are necessary relations). So the belief that 2+2=4 is not physical.

Now some analytic functionalists will register the following objection to the argument we just gave:

> You say — and your argument essentially presupposes — that there is no entailment from *C is instantiated in R,* where is an object-independent microphysical concept, to *there is a belief that 2+2=4 in R.* But *there is* such an entailment. Given a knowledge of the kinematic and dynamic interrelation of the physical objects in R, it *could* be deduced, quite literally, whether there was a belief that 2+2=4 in R. Consider: let R be the region occupied by some computer. If you knew all the microphysical facts in R — i.e. if you knew just what microphysical concepts were instantiated in R — then you would have as fine-grained a knowledge as possible of the character and organization of the states of affairs obtaining in R. You would, in effect, know everything there was to know about the distribution of mass-energy in R — about the course and intensity of electric currents, the mechanical interactions, and so forth. But it seems to me that, on the basis of this knowledge, you *could* deduce that something in R believed that 2+2=4. You would know that the computer generated such and such and output in response to thus and such input, and you would know the intervening electrical and mechanical facts. Now if we are functionalists about the belief that 2+2=4 — that is, if we say that x qualifies as such a belief *wholly* in virtue of its causal liaisons — then, on the basis of the aforementioned physical facts, one *could* deduce that the computer believed that 2+2=4. So *there would be* an entailment from *CI, C2...Cn are instantiated in R* — where C1, C2...Cn are microphysical concepts — to *there is a belief that 2+2=4 in R.*[22]

This seems to me to involve a very wrong-headed and overly reductive conception of belief. But I cannot pursue that here. Nonetheless, what the analytic functionalist says about the belief that 2+2=4 has virtually no bearing on examples involving concept s of phenomenally pregnant states — for such states are not plausibly regarded functionally. (It is exceedingly implausible to say that the essence of being a pain or a ticklish sensation is having certain causes or effects.) And we can use this fact to circumvent the objector's point.

Consider the concept *is a pain.* For reasons we've seen, if materialism is right, then the concept *is a pain is* instantiated in R wholly in virtue of the fact that some object-independent microphysical concept C is instantiated in R. Now *is a pain* is object-independent. Suppose that Bob and twin-Bob are exactly alike in respect of their mental states, narrowly individuated (i.e. considered apart from any environmental facts). And suppose that Bob has the concept *is a pain (i.e.* he knows what it is to attribute pain to someone or something). Under those circumstances, could it possibly be *denied* that twin-Bob had the concept *is a pain?* Surely not. Surely if Bob knows what a pain is, then so does twin-Bob: facts about the environmental causes of Bob's and twin-Bob's mental contents are totally irrelevant. If Bob has the concept *is a pain,* and twin-Bob is his exact duplicate — in all respects *modulo* those having to do with the environmental causes of his mental contents — then if Bob has the concept is *a pain,* so does twin-Bob.

So *is a pain is* an object-independent concept. And, so by assumption, is C. But surely there is no *entailment* from *C is instantiated in R* to *the concept* is a pain *is instantiated in R.* For any object-independent microphysical concepts C1, C2...Cn, from the fact that C1, C2...Cn are instantiated in R, it - surely

22 The classical statement of analytic functionalism is found in Lewis [1].

cepts C1, C2...Cn, from the fact that C1, C2...Cn are instantiated in R, it surely does not logically follow that *is a pain* is instantiated in R; there is no *entailment* from *Cl, C2...Cn are instantiated in R* to *the concept is* a pain is *instantiated in R.* In other words, there is no *entailment* from *Cl, C2...Cn are instantiated in R* to *there is a pain in R.* Now since C1, C2...Cn and *is a pain* are object-independent concepts, necessary relations that hold between the former and the latter are *entailment* (basically, logical) relations. So if, from the fact that C1, C2...Cn were instantiated, it were really *necessary* that *is a pain* be instantiated, there would be an *entailment* from *C1, C2...Cn are instantiated in R* to *the concept* is a pain *is instantiated in R (i.e. there is a pain in R).* But there is no entailment; so there is no necessary connection between the concepts *C1, C2...Cn* being instantiated in R, on the one hand, and *is a pain* being in instantiated in R, on the other. (That is, there is no necessary connection between *C1, C2... Cn* being instantiated in R, on the one hand, and there being a pain in R, on the other.) Now C1, C2...Cn stand for *any* object-independent microphysical concepts one might choose. *So for any* object-independent microphysical concepts one might choose, there is no necessary relation between those concepts being instantiated in R, on the one hand, and there being in a pain in R, on the other. For a microphysical state of affairs to obtain in R just is for some microphysical concept to be instantiated in R.

So, it follows that *for any* microphysical state of affairs S that can be described in object-independent concepts, S's obtaining in R does *not* necessitate there being a pain in R. From this, of course, it follows that *no* such micro-physical state of affairs (in R, or any other region) necessitates the occurrence of some pain in that region. Therefore, pain is not identical with, or supervenient upon, the occurrence of any microphysical state of affairs or, therefore, any physical state of affairs.

Once it is granted that materialism is false, then in order to reconcile the causal integratedness of the mental and the physical with the fact that the physical world is causally self-contained, we must adopt the strange and counter-intuitive, but otherwise (as far as I can tell) unexceptionable, view advocated in this paper.

Bibliography

CHALMERS, David J. *The Conscious Mind.* Oxford University Press, 1996.
KAPLAN, David [1] `Demonstratives'. In *:Themes From Kaplan.* Ed. Joseph Almog, John Perry, and Howard Wettstein. Oxford University Press, 1989.
KIM, Jaegwon. [1] *Philosophy of Mind.* Westview Press, 1996.

KRIPKE, Saul. [1] *Naming and Necessity.* Harvard University Press, 1980.
_____. [2] `Identity and Necessity'. In: *Metaphysics:* An Anthology. Ed. Jaegwon Kim and Ernest Sosa. Blackwell Publishers Ltd., 1999.
LEWIS, David. [1] "Psychophysical and theoretical identifications". *Australasian Journal of Philosophy,* 50 p, p.249-258.
McGINN, Colin. [1] *The Character of Mind.* Oxford University Press, 1997. NAGEL, Ernest. [1] *The Structure of Science.* Hackett, 1979.
PUTNAM, Hilary. [1] "The Meaning of `Meaning'." In: Putnam's *Philosophical Papers II:* Mind, Language, and Reality. Cambridge University Press, 1975.
RUSSELL, Bertrand Arthur William. [1] *Human Knowledge:* Its Scope and Limits. Routledge, 1992.
_____ . *[2] The Analysis of Mind.* George Allen & Unwinn Ltd., 1954.
SEARLE, John R. [1] *Mind, Language, and Society.* Weigenfeld and Nicolson, 1999.
_____ . *[2] The Rediscovery of the Mind.* The MIT Press, 1995.
SMART, J. J.C. [1] `Sensations and Brain Processes'. Reprinted In: *The Mind.* Ed. Daniel N. Robinson. Oxford University Press, 1998.

Replicated from **Kriterion**, Belo Horizonte, v.45, n.109, p.81-135, Jan./June 2004.