

Nicholas Humphrey, 2000, "How to solve the mind-body problem," *Journal of Consciousness Studies*, 7, 5-20. [Commentaries and my reply appear in the same issue of the Journal].

HOW TO SOLVE THE MIND-BODY PROBLEM

Two hundred and fifty years ago Denis Diderot, commenting on what makes a great natural philosopher, wrote:

They have watched the operations of nature so often and so closely that they are able to guess what course she is likely to take, and that with a fair degree of accuracy, even when they take it into their heads to provoke her with the most outlandish experiments. So that the most important service they can render to [others] . . . is to pass on to them that spirit of divination by means of which it is possible to *smell out*, so to speak, methods that are still to be discovered, new experiments, unknown results.

Whether Diderot would have claimed such a faculty in his own case is not made clear. But I think there is no question we should claim it for him. For, again and again, Diderot made astonishingly prescient comments about the future course of natural science. Not least, this:

Just as in mathematics, all the properties of a curve turn out upon examination to be all the same property, but seen from different aspects, so in nature, when experimental science is more advanced, we shall come to see that all phenomena, whether of weight, elasticity, attraction, magnetism or electricity, are all merely aspects of a single state.

Admittedly the grand unifying theory that Diderot looked forward to has not yet been constructed. And contemporary physicists are still uncertain whether such a theory of *everything* is possible even in principle. But, within the narrower field that constitutes the study of *mind and brain*, cognitive scientists are increasingly confident of its being possible to have a unifying theory of these *two* things.

They — we — all assume that the human mind and brain are, as Diderot anticipated, aspects of a single state — a single state, in fact, of the material world, which could in principle be fully described in terms of its microphysical components. We assume that each and every instance of a human mental state is *identical* to a brain state, **mental state, m = brain state, b** , meaning that the mental state and the brain state pick out the same thing at this microphysical level. And usually we further assume that the nature of this identity is such that

each type of mental state is multiply realisable, meaning that instances of this one type can be identical to instances of several different types of brain states that happen to be functionally equivalent.

What's more, we have reason to be confident that these assumptions are factually correct. For, as experimental science grows more advanced, we are indeed *coming to see* that mind and brain are merely aspects of a single state. In particular, brain-imaging studies, appearing almost daily in the scientific journals, demonstrate in ever more detail how specific kinds of mental activity (as reported by a mindful subject) are precisely correlated with specific patterns of brain activity (as recorded by external instruments). *This* bit of the brain lights up when a man is in pain, *this* when he conjures up a visual image, *this* when he tries to remember which day of the week it is, and so on.

No doubt many of us would say we have known all along that such correspondences must in principle exist. So that our faith in mind-brain identity hardly needs these technical demonstrations. Even so, it is, to say the least, both satisfying and reassuring to see the statistical facts of the identity being established, as it were, right before our eyes.

Yet it's one thing to see *that* mind and brain are aspects of a single state, but quite another to see *why* they are. It's one thing to be convinced by the statistics, but another to understand — as surely we all eventually want to — the causal or logical principles involved. Even while we have all the evidence required for *inductive generalisation*, we may still have no basis for *deductive explanation*.

Let's suppose, by analogy, that we were to come to see, through a series of "atmospheric-imaging" experiments, that whenever there is a visible shaft of lightning in the air there is a corresponding electrical discharge. We might soon be confident that the lightning and the electrical discharge are aspects of one and the same thing, and we should certainly be able to predict the occurrence of lightning whenever there is the electrical discharge. Even so, we might still have not a clue about what *makes* an electrical discharge manifest also as lightning.

Likewise, we might one day have collected so much detailed information about mind-brain correlations that we can predict which mental state will supervene on any specific brain state. Even so we might still have no idea as to the reasons why this brain state yields this mental state, and hence no way of deducing one from the other *a priori*.

But with lightning there could be — and of course historically there was — a way to progress to the next stage. The physico-chemical causes that underlie the identity could be discovered through further experimental research and new theorising. Now the question is whether the same strategy will work for mind and brain.

When experimental science is *even more* advanced, shall we come to see not only *that* mind and brain are merely aspects of a single state, but *why* they have to be so? Indeed, shall we be able to see how an identity that might otherwise appear to be mysteriously contingent is in fact transparently necessary?

A few philosophers believe the answer must be No. Or, at any rate, they believe we shall never achieve this level of understanding for every single feature of the mind and brain. They would point out that not all identities are in fact open to analysis in logical or causal terms, even in principle. Some identities are metaphysically primitive, and have simply to be taken as givens. And quite possibly some basic features of the mind are in this class. David Chalmers, for example, takes this stance when he argues for a version of epiphenomenal dualism in which consciousness just happens to be a fundamental, non-derivative, property³ of matter.

But even supposing — as most people do — that all the *interesting* identities are in fact analyzable in principle, it might still be argued that not all of them will be open to analysis by us human beings. Thus Colin McGinn believes that the reason why a full understanding of the mind-brain identity will never be achieved is not because the task is logically impossible but because there are certain kinds of understanding — and this is clearly one of them — which must for ever lie beyond our intellectual reach: no matter how much more factual knowledge we accumulate⁴ about mind and brain, we simply do not have what it would take to come up with the right theory.

The poet Goethe, much earlier, counseled against what he considered to be the hubris of our believing that we humans can in fact solve every problem. “In Nature,” he said, “there is an accessible element and an inaccessible. . .

Anyone who does not appreciate this distinction may wrestle with the inaccessible for a lifetime without ever coming near to the truth. He who does recognise it and is sensible will keep to the accessible and by progress in every direction within a field and consolidation, may even be able to wrest something from the inaccessible along the way — though here he will in the end have to admit that some things can only be grasped up to a certain point, and that Nature always retains behind her something⁵ problematic which it is impossible to fathom with our inadequate human faculties.

It is not yet clear how far — if at all - such warnings should be taken seriously. Diderot, for one, would have advised us to ignore them. Indeed Diderot, ever the scientific modernist, regarded any claim by philosophers to have found limits to our understanding, and

thus to set up No-Go areas, as an invitation to science (or experimental philosophy) to prove such rationalist philosophy wrong.

Experimental philosophy knows neither what will come nor what will not come out of its labours; but it works on without relaxing. The philosophy based on reasoning, on the contrary, weighs possibilities, makes a pronouncement and stops short. It boldly said: "light cannot be decomposed": experimental philosophy heard, and held its tongue in its presence for whole centuries; then suddenly it produced the prism, and said, "light can be decomposed".⁶

The hope now of cognitive scientists is of course that there is a prism awaiting discovery that will do for the mind-brain identity what Newton's prism did for light - a prism that will again send the philosophical doubters packing.

I am with them in this hope. But I am also very sure we shall be making a mistake if we ignore the philosophical warnings entirely. For there is no question that the likes of McGinn and Goethe might have a point. Indeed, I'd say they might have more than a point: they will actually become right by default, *unless and until we can set out the identity in a way that meets certain minimum standards for explanatory possibility*.

To be precise, we need to recognise that there can be no hope of scientific progress so long as we continue to write down the identity in such a way that the mind terms and the brain terms are patently *incommensurable*.⁷ The problem will be especially obvious if the *dimensions* do not match up.

I use the word "dimensions" here advisedly. When we do physics at school we are taught that the "physical dimensions" of each side of an equation must be the same. If one side has the dimensions of a volume, the other side must be a volume too, and it cannot be, for example, an acceleration; if one side has the dimensions of power, the other side must be power too and it cannot be momentum; and so on. As A. S. Ramsey put this in his classical *Dynamics* textbook: "The consideration of dimensions is a useful check in dynamical work, for each side of an equation must represent the same physical thing and therefore must be of the same dimensions in mass [m], space [s] and time [t]".⁸

Indeed so strong a constraint is this that, as Ramsey went on, "sometimes a consideration of dimensions alone is sufficient to determine the form of the answer to a problem." For example, suppose we want to know the form of the equation that relates the energy contained in a lump of matter, **E**, to its mass, **M**, and the velocity of light, **C**. Since **E** can only have the dimension $\text{ms}^2 \text{t}^{-2}$, **M** the dimension **m** and **C** the dimension st^{-1} , we can

conclude without further ado that the equation must have the form $\mathbf{E} = \mathbf{MC}^2$. By the same token, if anyone were to propose instead that $\mathbf{E} = \mathbf{MC}^3$, we would know immediately that something was wrong.

But what is true of these dynamical equations is of course just as true of all other kinds of identity equations. We can be sure in advance that, if any proposed identity is to have even a chance of being valid, both sides must represent the same *kind* of thing. Indeed we can generalise this beyond physical dimensions, to say that both sides must have the same conceptual dimensions, which is to say they must belong to the same generic class.

So, if it is suggested for example that Mark Twain and Samuel Clemens are identical, **Mark Twain = Samuel Clemens**, we can believe it because both sides of the equation are in fact people. Or, if it is suggested that Midsummer Day and 21st June are identical, **Midsummer Day = 21st June**, we can believe it because both sides are days of the year. But were someone to suggest that Mark Twain and Midsummer Day are identical, **Mark Twain = Midsummer Day**, we should know immediately this equation is a false one.

Now, to return to the mind-brain identity: when the proposal is that a certain mental state is identical to a certain brain state, **mental state, m = brain state, b** , the question is: do the dimensions of the two sides match?

The answer surely is, Yes, sometimes they do, or at any rate they can be made to.

Provided cognitive science delivers on its promise, it should soon be possible to characterise many mental states in computational or functional terms, i.e. in terms of rules connecting inputs to outputs. But brain states too can relatively easily be described in these same terms. So it should then be quite straightforward, in principle, to get the two sides of the equation to line up.

Most of the states of interest to psychologists — states of remembering, perceiving, wanting, talking, thinking, and so on — are in fact likely to be amenable to this kind of functional analysis. So, although it is true there is still a long way to go before we can claim much success in practice, at least the research strategy is clear.

We do an experiment, say, in which we get subjects to recall what day of the week it is, and at the same time we record their brain activity by MRI. We discover that whenever a person thinks to himself “today is Tuesday”, a particular area of the brain lights up. We postulate the identity: **recalling that today is Tuesday = activity of neurons in the calendula nucleus.**

We then try, on the one hand, to provide a computational account of what is involved in this act of recalling the day; and, on the other hand, we examine the local brain activity and

try to work out just what is being computed. Hopefully, when the results are in, it all matches nicely. A clear case of **Mark Twain = Samuel Clemens**.

But of course cases like this are notoriously the “easy” cases — and they are not the ones that most philosophers are really fussed about. The “hard” cases are precisely those where it seems that this kind of functional analysis is not likely to be possible. And this means especially those cases that involve *phenomenal consciousness*: the subjective sensation of redness, the taste of cheese, the pain of a headache, and so on. These are the mental states that Isaac Newton dubbed sensory “phantasms”,⁹ and which are now more generally (although often less appropriately) spoken of as “qualia”.

The difficulty in these latter cases is not that we cannot establish the factual evidence for the identity. Indeed this part of the task may be just as easy as in the case of cognitive states such as remembering the day. We do an experiment, say, in which we get subjects to experience colour sensations, while again we examine their brain by MRI. We discover that whenever someone has a red sensation, there is activity in cortical area Q6. So we postulate the identity: **phantasm of red = activity in Q6 cortex**.

So far, so good. But it is the next step that is problematical. For now, if we try the same strategy as before and attempt to provide a functional description of the phantasm so as to be able to match it with a functional description of the brain state, the way is barred. No one it seems has the least idea how to characterise the phenomenal experience of redness in functional terms — or for that matter how to do it for any other variety of sensory phantasm. And in fact there are well-known arguments (such as the Inverted Spectrum) that purport to prove that it cannot be done, even in principle.

If not a functional description, then, might there be some other way of describing these elusive states, which being also applicable to brain states, could save the day? Unfortunately, the philosophical consensus seems to be that the answer must be No. For many philosophers seem to be persuaded that phenomenal states and brain states are indeed essentially such different kinds of entity that there is simply no room whatever for negotiation. Colin McGinn, in a fantasy dialogue, expressed the plain hopelessness of it sharply: “Isn’t it perfectly evident to you that . . . [the brain] is just the wrong kind of thing to give birth to [phenomenal] consciousness. You might as well assert that numbers emerge from biscuits or ethics from rhubarb”.¹⁰ A case of **Mark Twain = Midsummer Day**.

Yet, as we’ve seen, this will not do! At least not if we are still looking for explanatory understanding. So, where are we scientists to turn?

Let's focus on the candidate identity: **phantasm, p = brain state, b** . Given that the statistical evidence supporting it remains as strong as ever, there would seem to be three ways that we can go.

1. We can accept that, despite everything, this equation is in fact false. Whatever the statistical evidence for there being a correlation between the two, there is really *not* an identity between the two states. Indeed all the correlation shows is just that: that the states are *co-related*. And if we want to pursue it, we shall have then to go off and look for some other theoretical explanation for this correlation — God's whim, for instance. (This would have been Descartes' preferred solution).

2. We can continue to believe in the equation, while at the same time we grudgingly acknowledge that we have met our match: either the identity does not have an explanation or else the explanation really is beyond our human reach. And, recognising now that there is no point in pursuing it, we shall be able, with good conscience, to retire and do something else. (This is McGinn's preferred solution).

3. We can doggedly insist both that the identity is real and that we shall explain it somehow — when eventually we do find the way of bringing the dimensions into line. But then, despite the apparent barriers, we shall have to set to work to brow-beat the terms on one side or other of the identity equation in such way as to *make* them line up. (This is my own and I hope a good many others' preferred solution).

Now, if we do choose this third option, there are several possibilities.

One strategy would be to find a new way of conceiving of sensory phantasms so as to make them more obviously akin to brain states. But, let's be careful. We must not be *too* radical in redefining these phantasms or we shall be accused of redefining away the essential point. Daniel Dennett's sallies in this direction can be a warning to us.¹¹ His suggestion that sensations are nothing other than complex behavioural (even purely linguistic?) dispositions, while defensible in his own terms, has proved too far removed from most people's intuitions to be persuasive.

An alternative strategy would be to find a new way of conceiving of brain states so as to make them more like sensory phantasms. But again we must not go too far. Roger Penrose is the offender this time.¹² His speculations about the brain as a quantum computer, however ingenious, have seemed to most neuroscientists to require too much special pleading to be taken seriously.

Or then again, there would be the option of doing *both*. My own view is that we should indeed try to meddle with both sides of the equation to bring them into line. Dennett expects all the compromise to come from the behavioural psychology of sensation, Penrose expects it all to come from the physics of brain states. Neither of these strategies seems likely to deliver what we want. But it's amazing how much more promising things look when we allow some give on *both* sides — when we attempt to adjust our concept of sensory phantasms *and* our concept of brain states until they do match up.

So, *this*, I suppose, is how to solve the mind-brain problem. We shall need to work on both sides to define the relevant mental states and brain states in terms of concepts that really do have *dual currency* — being equally applicable to the mental and the material. And now all that remains, for this paper, is to do it.

Then, let's begin. **phantasm,*p*** = **brain state, *b***. Newton himself wrote: “To determine . . . by what modes or actions light produceth in our minds the phantasms of colours is not so easy. And I shal not mingle conjectures with certainties”.¹³ Three and a half centuries later, let us see if we can at least mix some certainties with the conjectures.

First, on one side of the equation, these sensory phantasms. Precisely what are we are talking about here? What kind of thing are they? What indeed are their dimensions?

Philosophers are — or at any rate have become in recent years — remarkably cavalier about the need for careful definition in this area. They bandy about terms such as “phenomenal properties,” “what it's like,” “conscious feelings,” and so on, to refer to whatever it is that is at issue when people point inwardly to their sensory experience — as if the hard-won lessons of positivist philosophy had never been learned. In particular that over-worked term “qualia,” which did at least once have the merit of meaning something precise (even if possibly vacuous¹⁴), is now widely used as a catch-all term for anything vaguely subjective and qualitative.

It is no wonder, then, that working scientists, having been abandoned by those who might have been their pilots, have tended to lose their way even more comprehensively. Francis Crick and Christoph Koch, for example, begin a recent paper by saying that “everyone has a rough idea of what is meant by consciousness” and that “it is better to avoid a precise definition of consciousness”.¹⁵ In the same vein Susan Greenfield, writes “consciousness is impossible to define . . . perhaps then it is simply best to give a hazy description, something like consciousness being ‘your first-person, personal world’”.¹⁶ While Antonio Damasio is fuzzier still: “Quite candidly, this first problem of consciousness is the problem of how we get a

‘movie in the brain’ . . . the fundamental components of the images in the movie metaphor are thus made of qualia.”¹⁷

But this is bad. Hazy or imprecise descriptions can only be a recipe for trouble. And, anyway, they are unnecessary. For the fact is we have for a long time had the conceptual tools for seeing through the haze and distinguishing the phenomenon of central interest.

Try this. Look at a red screen, and consider what mental states you are experiencing. Now let the screen suddenly turn blue, and notice how things change. The important point to note is that there are *two* quite distinct parts to the experience, and *two* things that change.

First (and I mean first), there is a change in the experience of something happening to yourself — the bodily sensation of the quality of light arriving at your eye. Second, there is a change in your attitude towards something in the outer world — your perception of the colour of an external object.

It was Thomas Reid, genius of the Scottish enlightenment, who over two hundred years ago first drew philosophical attention to the remarkable fact that we human beings — and presumably many other animals also — do in fact use our senses in these two quite different ways:

The external senses have a double province — to make us feel, and to make us perceive. They furnish us with a variety of sensations, some pleasant, others painful, and others indifferent; at the same time they give us a conception and an invincible belief of the existence of external objects. . .

Sensation, taken by itself, implies neither the conception nor belief of any external object. It supposes a sentient being, and a certain manner in which that being is affected; but it supposes no more. Perception implies a conviction and belief of something external - something different both from the mind that perceives, and the act of perception. Things so different in their nature ought to be distinguished.¹⁸

For example, Reid said, we smell a rose, and two separate and parallel things happen: we both feel the sweet smell at our own nostrils and we perceive the external presence of a rose. Or, again, we hear a hooter blowing from the valley below: we both feel the booming sound at our own ears and we perceive the external presence of a ship down in the Firth. In general we can and usually do use the evidence of sensory stimulation *both* to provide a “subject-centred affect-laden representation of what’s happening to me”, *and*¹⁹ to provide “an objective, affectively neutral representation of what’s happening out there”.

Now it seems quite clear that what we are after when we try to distinguish and define the realm of sensory phantasms is the first of these: sensation rather than perception. Yet one reason why we find it so hard to do the job properly is that it is so easy to muddle the two up. Reid again:

[Yet] the perception and its corresponding sensation are produced at the same time. In our experience we never find them disjoined. Hence, we are led to consider them as one thing, to give them one name, and to confound their different attributes. It becomes very difficult to separate them in thought, to attend to each by itself, and to attribute nothing to it which belongs to the other. To do this, requires a degree of attention to what passes in our own minds, and a talent for distinguishing things that differ, which is not to be expected in the vulgar, and is even rarely found in philosophers. . .

I shall conclude this chapter by observing that, as the confounding our sensations with that perception of external objects which is constantly conjoined with them, has been the occasion of most of the errors and false theories of philosophers with regard to the senses; so the distinguishing²⁰ these operations seems to me to be the key that leads to a right understanding of both.

To repeat: sensation has to do with the self, with bodily stimulation, with feelings about what's happening *now to me* and how *I* feel about it; perception by contrast has to do with judgements about the objective facts of the external world. Things so different in their nature *ought* to be distinguished. Yet rarely are they. Indeed many people still assume that perceptual judgements and even beliefs, desires and thoughts can have a pseudo-sensory phenomenology in their own right.

Philosophers will be found claiming for example that “there is something it is like” not only to have sensations such as feeling warmth on one's skin, but also to have perceptions such as seeing the shape of a distant cube, and even to hold propositional attitudes such as believing that Paris is the capital of France.²¹ Meanwhile psychologists, adopting a half-understood vocabulary borrowed from philosophy, talk all too casually about such hybrid notions as the perception of “dog qualia” on looking at a picture of a dog.²² While these category mistakes persist we might as well give up.

So this must be the first step: we have to mark off the phenomenon that interests us — sensation — and get the boundary *in the right place*. But then the real work of analysis begins. For we must home in on what *kind of thing* we are dealing with.

Look at the red screen. You feel the red sensation. You perceive the red screen. We do in fact talk of both sensation and perception in structurally similar ways. We talk of *feeling* or *having* sensations – as if somehow these sensations, like perceptions, were the *objects* of our sensing, sense *data*, out there waiting for us to grasp them or observe them with our mind's eye.

But, as Reid long ago recognised, our language misleads us here. In truth, sensations are no more the objects of sensing than, say, volitions are the objects of willing, intentions the objects of intending, or thoughts the object of thinking.

Thus, *I feel a pain; I see a tree*: the first denoteth a sensation, the last a perception. The grammatical analysis of both expressions is the same: for both consist of an active verb and an object. But, if we attend to the things signified by these expressions, we shall find that, in the first, the distinction between the act and the object is not real but grammatical; in the second, the distinction is not only grammatical but real.

The form of the expression, *I feel pain*, might seem to imply that the feeling is something distinct from the pain felt; yet in reality, there is no distinction. As *thinking a thought* is an expression which could signify no more than *thinking*, *so feeling a pain* signifies no more than *being pained*. What we have said of pain is applicable to every other mere sensation.²³

So sensory awareness is an *activity*. We do not *have* pains *we get to be* pained.

This is an extraordinarily sophisticated insight. And all the more remarkable that Reid should have come to it two hundred years before Wittgenstein was tearing his hair about similar problems and not getting noticeably further forward.

Even so, I believe Reid himself got only part way to the truth here. For my own view (developed in detail in my book, *A History of the Mind*²⁴) is that the right expression is not so much “being pained” as “paining”. That is to say, sensing is not a passive state at all, but rather a form of active engagement with the stimulus occurring at the body surface.

When, for example, I feel pain in my hand, or taste salt on my tongue, or equally when I have a red sensation at my eye, I am not *being* pained, or *being* stimulated saltily, or *being* stimulated redly. In each case I am in fact the active agent. I am not sitting there passively absorbing what comes *in from* the body surface, I am reflexly reaching out *to* the body surface

with an evaluative response — a response appropriate to the stimulus and the body part affected.

Furthermore, it is this *efferent activity* that I am aware of. So that what I actually experience as the feeling — the sensation of what is happening to me — is my reading of my own response to it. Hence the quality of the experience, the way it feels, instead of revealing the way something is being done to me, reveals the very way something is being done by me.

This is how I feel about what's happening right now at my hand — I'm feeling painfully about it!

———*This* is how I feel about what's happening right now at this part of the field of my eye — I'm feeling redly about it!

In my book I proposed that we should call the activity of sensing “sentition”. The term has not caught on. But I bring it up again here, in passing, because I believe we can well do with a word that captures the active nature of sensation: and sentition, resonating as it does with volition and cognition, sounds the right note of directed self-involvement.

The idea, to say it again, is that this sentition involves the subject “reaching out to the body surface with an evaluative response — a response appropriate to the stimulus and the body part affected.” This should not of course be taken to imply that such sensory responses actually result in overt bodily behaviour — at least certainly not in human beings as we are now. Nonetheless I think there is good reason to suppose that the responses we make today have in fact *evolved* from responses that in the past did carry through into actual behaviour. And the result is that even today the experience of sensation retains many of the original characteristics of the experience of true bodily action.

Let's consider, for example, the following five defining properties of the experience of sensation — and, in each case, let's compare an example of sensing, *feeling a pain in my hand*, with an example of bodily action, *performing a hand wave*.

1. Ownership. Sensation always *belongs to the subject*. When I have the pain my hand, *I own* the paining, it's mine and no one else's, I am the one and only *author* of it . . . as when I wave my hand, *I own and am the author of* the action of waving.

2. Bodily location. Sensation is always *indexical and invokes a particular part of the subject's body*. When I have the pain in my hand, the paining intrinsically involves *this* part of *me* . . . as when I wave my hand the waving too intrinsically involves *this* part of *me*.

3. Presentness. Sensation is always *present tense, ongoing and imperfect*. When I have the pain in my hand, the paining is in existence just *now for the time being* . . . as when I wave my hand the waving too exists just *now*.

4. Qualitative modality. Sensation always has the feel of one of several *qualitatively distinct modalities*. When I have the pain in my hand, the paining belongs to the class of *somatic* sensations, quite different in their whole style from, say, the class of visual sensations or of olfactory ones . . . as when I wave my hand the waving belongs to the class of *hand-waves*, quite different in style from other classes of bodily actions such as, say, the class of face-smiles or of knee-jerks.

5. Phenomenal immediacy. Most important, sensation is always *phenomenally immediate*, and the four properties above are *self-disclosing*. Thus, when I have the pain in my hand my impression is simply that *my hand hurts*: and, when my hand hurts, the fact that it is *my hand* (rather than someone else's), that it is *my hand* (rather than some other bit of me), that it is hurting *now* (rather than some other time), and that it is acting in *a painful* fashion (rather than acting in a visual, gustatory or auditory fashion), are facts of which I am directly and immediately aware *for the very reason that it is I, the author of the paining, who make these facts* . . . just as when I wave my hand, my impression is simply that my hand waves, and all the corresponding properties of this action too are facts of which I, *the author of the wave*, am immediately aware for similar reasons.

Thus, in these ways, and others that I could point to, the positive analogies between sensations and bodily activities add up. And yet, I acknowledge right away that there is also an obvious disanalogy: namely that, to revert to that old phrase, it is “like something” to have sensations, but not like anything much to engage in most other bodily activities!

To say the least, our experience of other bodily activities is usually very much shallower. When I wave my hand there may be, perhaps, the ghost of some phenomenal experience. But surely what it's like to wave hardly compares with what it's like to feel pain, or taste salt or sense red. The bodily activity comes across as a flat and papery phenomenon, whereas the sensation seems so much more velvety and thick. The bodily activity is like an unvoiced whisper, whereas the sensation is like the rich *self-confirming* sound of a piano with the sustaining pedal down.

Of course neither metaphor quite captures the difference in quality I am alluding to. But still I think the sustaining pedal brings us surprisingly close. For I believe that ultimately the key to an experience being “like something” does in fact lie in the experience *being like*

itself in time - hence *being about itself*, or *taking itself as its own intentional object*. And this is achieved, in the special case of sensory responses, through a kind of *self-resonance* that effectively stretches out the present moment to create what I have called the *thick moment of consciousness*.²⁵

There are, of course, loose ends to this analysis and ambiguities. But I'd say there are surely fewer of both than we began with. And this is the time to take stock, and move on.

The task was to recast the terms on each side of the mind-brain identity equation, **phantasm, *p* = brain state, *b***, so as to make them look more like each other.

What we have done so far is to redescribe the left hand side of the equation in progressively more concrete terms. Thus the phantasm of pain becomes the sensation of pain, the sensation of pain becomes the experience of actively paining, the activity of paining becomes the activity of reaching out to the body surface in a painful way, and this activity becomes self-resonant and thick. . . And with each step we have surely come a little closer to specifying something of a *kind* that we can get a handle on.

We can therefore turn our attention to the right hand side of the equation. As Ramsey wrote, "Sometimes a consideration of dimensions alone is sufficient to determine the form of the answer to a problem." If we now have this kind of thing on the mind side, we need to discover something like it on the brain side. If the mind term involves a state of *actively doing something about something*, namely issuing commands for an evaluative response addressed to body surface, then the brain term must also be a state of actively doing something about something, presumably doing the corresponding thing. If the mind term involves *self-resonance*, then the brain state must also involve self-resonance. And so on.

Is this still the impossibly tall-order that it seemed to be earlier — still a case of ethics on one side, rhabarb on the other? No, I submit that the hard problem has in fact been transformed into a relatively easy problem. For we are now dealing with something on the mind side that surely *could* have the same dimensions as a brain state *could*. Concepts such as "indexicality," "present-tenseness," "modal quality," and "authorship" are indeed dual currency concepts of just the kind required.

It looks surprisingly good. We can surely now imagine what it would take on the brain side to make the identity work. But I think there is double cause to be optimistic. For, as it turns out, this picture of what is needed on the brain side ties in beautifully with a plausible account of the evolution of sensations.

I shall round off this paper by sketching in this evolutionary history. And if I do it in what amounts to cartoon form, I trust this will at least be sufficient to let the major themes come through.

Let's return, then, in imagination to the earliest of times and imagine a primitive amoeba-like animal floating in the ancient seas.

This animal has a defining edge to it, a structural boundary. This boundary is crucial: the animal *exists* within this boundary – everything within it is part of the animal, belongs to it, is part of "self", everything outside it is part of "other". The boundary holds the animal's own substance in and the rest of the world out. The boundary is the vital frontier across which exchanges of material and energy and information can take place.

Now light falls on the animal, objects bump into it, pressure waves press against it, chemicals stick to it. No doubt some of these surface events are going to be a good thing for the animal, others bad. If it is to survive it must evolve the ability to sort out the good from the bad and to respond differently to them - reacting to this stimulus with an ow! to that with an ouch! to this with a whowee!

Thus, when, say, salt arrives at its skin it detects it and makes a characteristic wriggle of activity — it wriggles saltily. When red light falls on it, it makes a different kind of wriggle — it wriggles redly. These are adaptive responses, selected because they are appropriate to the animal's particular needs. Wriggling saltily has been selected as the best response to salt, while wriggling sugarly, for example, would be the best response to sugar. Wriggling redly has been selected as the best response to red light, while wriggling bluely would be the best response to blue light.

To begin with these wriggles are entirely local responses, organised immediately around the site of stimulation. But later there develops something more like a reflex arc passing via a central ganglion or proto-brain: information arrives from the skin, it gets assessed, and appropriate adaptive action is taken.

Still, as yet, these sensory responses are nothing other than responses, and there is no reason to suppose that the animal is in any way mentally aware of what is happening. Let's imagine however that, as this animal's life becomes more complex, the time comes when it will indeed be advantageous for it to have some kind of inner knowledge of what is affecting it, which it can begin to use as a basis for more sophisticated planning and decision making. So it needs the capacity to form *mental representations* of the sensory stimulation at the surface of its body and how it feels about it.

Now, one way of developing this capacity might be to start over again with a completely fresh analysis of the incoming information from the sense organs. But this would be to miss a trick. For, the fact is that all the requisite details about the stimulation — where

the stimulus is occurring, what kind of stimulus it is, and how it should be dealt with — are already encoded in the command signals the animal is issuing when it makes the appropriate sensory response.

Hence, all the animal needs to do to represent the stimulation is to pick up on these already-occurring command signals. For example, to sense the presence of salt at a certain location on its skin, it need only monitor its own signals for wriggling saltily at that location, or, equally, to sense the presence of red light it need only monitor its signals for wriggling redly.

Note well, however, that all this time the animal's concern is merely with what's occurring at its body surface. By monitoring its own responses, it forms a representation of "WHAT IS HAPPENING TO ME". But, at this stage, the animal neither knows nor cares *where the stimulation comes from, let alone what the stimulation may imply about the world beyond its body.*

Yet wouldn't it be better off if it *were* to care about the world beyond? Let's say a pressure wave presses against its side . . . wouldn't it be better off if, besides being aware of feeling the pressure wave as such, it were able to interpret this stimulus as signaling an approaching predator? A chemical odour drifts across its skin . . . wouldn't it be better off if it were able to interpret this stimulus as signaling the presence of a tasty worm? In short, wouldn't the animal be better off if, as well as reading the stimulation at its body surface merely in terms of its immediate affective value, it were able to interpret it as *a sign* of "WHAT IS HAPPENING OUT THERE"?

The answer of course is, Yes. And we can be sure that, early on, animals did in fact hit on the idea of using the information contained in body surface stimulation for this novel purpose - *perception* in addition to *sensation*. But the purpose was indeed *so* novel that it meant a very different style of information-processing was needed. When the question is "what is happening to me?", the answer that is wanted is qualitative, present-tense, transient, and subjective. When the question is "what is happening out there?", the answer that is wanted is quantitative, analytical, permanent, and objective.

So, to cut a long story short, there developed in consequence two parallel channels to subserve the very different readings we now make of an event at the surface of the body, sensation and perception: one providing an affect-laden modality-specific body-centred representation of what the stimulation is doing to me and how I feel about it, the other providing a more neutral, abstract, body-independent representation of the outside world.

Sensation and perception continued along relatively independent paths in evolution. But we need not be concerned further with perception in this paper. For it is the fate of sensation that matters to our narrative.

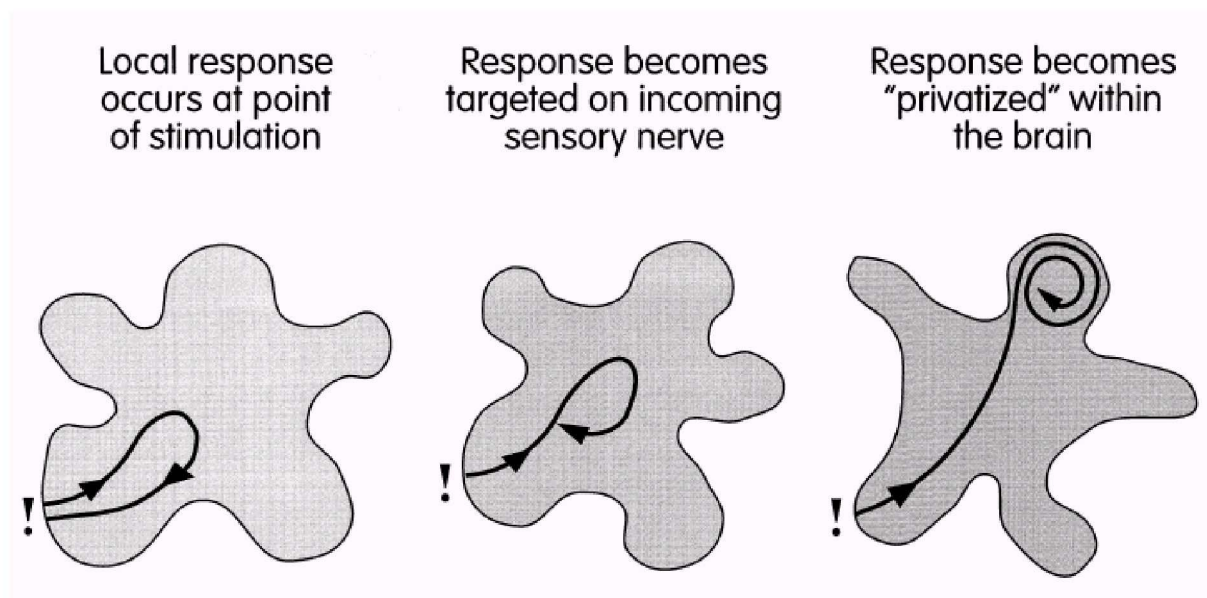
As we left it, the animal is actively responding to stimulation with public bodily activity, and its experience or proto-experience of sensation (if we can now call it that) arises from its monitoring its own command signals for these sensory responses. Significantly, these responses are still tied in to the animal's survival and their form is still being maintained by natural selection — and it follows that the form of the animal's sensory experience is also at this stage being determined in all its aspects by selection.

Yet, the story is by no means over. For, as this animal continues to evolve and to change its lifestyle, the nature of the selection pressures is bound to alter. In particular, as the animal becomes more independent of its immediate environment, it has less and less to gain from the responses it has always been making directly to the surface stimulus as such. In fact there comes a time when, for example, wriggling saltily or redly at the point of stimulation no longer has any adaptive value at all.

Then why not simply give up on this primitive kind of local responding altogether? The reason why not is that, even though the animal may no longer want to respond directly to the stimulation at its body surface as such, it still wants to be able to keep up to date mentally with what's occurring (not least because this level of sensory representation retains a crucial role in policing perception, see Chapter 10). So, even though the animal may no longer have any use for the sensory responses in themselves, it has by this time become quite dependent on the secondary representational functions that these responses have acquired. And since the way it has been getting these representations in the past has been by monitoring its own command signals for sensory responses, it clearly cannot afford to stop issuing these command signals entirely.

So, the situation now is this. In order to be able to represent “what's happening to me”, the animal must in fact continue to issue commands such as *would* produce an appropriate response at the right place on the body *if* they were to carry through into bodily behaviour. But, given that the behaviour is no longer wanted, it may be better if these commands remain virtual or as-if commands — in other words, commands which, while retaining their original intentional properties, do not in fact have any real effects.

The upshot is — or so I've argued — that, over evolutionary time, there is a slow but remarkable change. What happens is that the whole sensory activity gets “privatised”: the command signals for sensory responses get short-circuited before they reach the body surface, so that instead of reaching all the way out to the site of stimulation they now reach only to points closer and closer in on the incoming sensory nerve, until eventually the whole process becomes closed off from the outside world in an internal loop within the brain.



Now once *this* happens the role of natural selection must of course sharply diminish. The sensory responses have lost all their original biological importance and have in fact disappeared from view. Therefore selection is no longer involved in determining the form of these responses and *a fortiori* it can no longer be involved in determining the quality of the representations based on them.

But the fact is that this privacy has come about only at the very end, after natural selection has done its work to shape the sensory landscape. There is therefore every reason to suppose that the forms of sensory responses and the corresponding experiences have already been more or less permanently fixed. And although, once selection becomes irrelevant, these forms may be liable to drift somewhat, they are likely always to reflect their evolutionary pedigree. Thus responses that started their evolutionary life as dedicated wriggles of acceptance or rejection of a stimulus will still be recognisably of their kind right down to the present day.

Yet, something is not in place yet: the “thickness factor”. And, as it happens, there is a further remarkable evolutionary development to come — made possible by the progressive shortening of the sensory response pathway.

It has been true all along, ever since the days when sensory responses were indeed actual wriggles at the body surface, that they have been having *feedback* effects by modifying the very stimulation to which they are a response. In the early days, however, this feedback circuit was too round-about and slow to have had any interesting consequences. However, as and when the process becomes internalised and the circuit so much shortened, the conditions are there for a significant degree of recursive interaction to come into play. That's to say, the command signals for sensory responses begin to loop back upon themselves, becoming in the process partly self-creating and self-sustaining. These signals still *take their cue* from input from the body surface, and still get *styled* by it, but on another level they have become signals *about themselves*. To be the author of such recursive signals is to enter a new intentional domain.

To return to our identity equation: We *needed* a certain set of features on the brain side. We could have *invented* them if we were brave enough. But now, I submit, we actually have them *handed to us on a plate* by an evolutionary story that delivers on every important point.

I acknowledge that there is more to be done. And the final solution to the mind-body problem, if ever we do agree on it, may still look rather different from the way I'm telling it here. But the fact remains that this approach to the problem has to be the right one. There is no escaping the need for dual currency concepts – and any future theory will have to play by these rules.

Diderot wrote “A tolerably clever man began his book with these words: *Man, like all animals, is composed of two distinct substances, the soul and the body. If anyone denies this proposition it is not for him that I write.*’ I nearly shut the book. Oh! ridiculous writer, if I once admit these two distinct substances, you have nothing more to teach me.”²⁶

This paper has been about how to make one thing of these two.

1. Denis Diderot, 1754 / 1982, *On the Interpretation of Nature*, XXX, p.66, in *The Irresistible Diderot*, ed. J.H. Mason, London: Quartet.
2. Denis Diderot, 1754 / 1982, op. cit., XLV, p. 68.
3. David J. Chalmers, 1996, *The Conscious Mind*, Oxford: Oxford University Press.
4. Colin McGinn, 1989, “Can we solve the mind-body problem?”, *Mind*, 98, 349-366.
5. J.W. von Goethe, 1827, *Conversations with Eckermann*, 11th April 1827.

6. Denis Diderot, 1754 / 1982, op. cit., XXIII, p.46.
7. Compare the discussion of the same problem by J. Scott Kelso, 1995, *Dynamical Patterns: The self-Organisation of Brain and Behavior*, p. 29, Cambridge, Ma: MIT Press.
8. A.S. Ramsey, 1954, *Dynamics*, p. 42, Cambridge: Cambridge University Press.
9. Peter Munz, 1997, “The evolution of consciousness — silent neurons and the eloquent mind”, *Journal of Social and Evolutionary Systems*, 20, 7-28.
10. Colin McGinn, 1993, “Consciousness and cosmology: hyperdualism ventilated”, in *Consciousness*, ed. M. Davies & G.W. Humphreys, pp. 155-77, Oxford: Blackwell.
11. Daniel C. Dennett, 1991, *Consciousness Explained*, New York: Little Brown.
12. Roger Penrose, 1989, *The Emperor’s New Mind*, Oxford: Oxford University Press.
13. Isaac Newton, 1671, “A letter from Mr. Isaac Newton . . . containing his New Theory about Light and Colours”, p. 3085, *Philosophical Transactions of the Royal Society*, 80, 3075-87.
14. Daniel C. Dennett, 1988, “Quining Qualia”, in *Consciousness in Contemporary Science*, ed. A.J. Marcel & E. Bisiach, pp. 42-77, Oxford: Clarendon Press.
15. F. Crick and C. Koch, 2000, “The unconscious homunculus”, *Neuro-Psychoanalysis*, 2, 1-10.
16. Susan Greenfield, 1998, “How might the brain generate consciousness?”, p. 210, in *From Brains to Consciousness*, ed. Steven. Rose, pp. 210-227, London: Allen Lane.
17. Antonio Damasio, 2000, *The Feeling of What Happens*, p. 9, London: Heinemann.
18. Thomas Reid, 1785 / 1813, *Essays on the Intellectual Powers of Man*, ed. D. Stewart, Part II, Ch. 17 & 16, Charlestown: Samuel Etheridge.
19. Nicholas Humphrey, 1992, *A History of the Mind*, London: Chatto & Windus.
20. Thomas Reid, 1785 / 1813, op. cit. Part II, Ch. 17 & 16.
21. Ned Block, 1995, “On a confusion about a function of consciousness”, *Behavioral and Brain Sciences*, 18, 227-247; John R. Searle, 1992, *The Rediscovery of the Mind*, Cambridge Ma: MIT Press; see also my discussion in Nicholas Humphrey, 2000a, “Now you see it, now you don’t”, *Neuro-psychoanalysis*, 2, 14-17.

22. V.S. Ramachandran and W. Hirstein, 1997, “Three laws of qualia: what neurology tells us about the biological functions of consciousness”, *Journal of Consciousness Studies*, 4, 429-57.
23. Thomas Reid, 1764 / 1813, *An Inquiry into the Human Mind*, ed. D. Stewart, p. 112, Charlestown: Samuel Etheridge.
24. Nicholas Humphrey, 1992, op. cit.
25. Nicholas Humphrey, 1995, “The thick moment”, in *The Third Culture*, ed. J. Brockman, pp. 198-208, New York: Simon & Schuster.
26. Denis Diderot, 1774-80, *Elements of Physiology*, p. 139, in *Diderot: Interpreter of Nature*, trans. and ed. Jonathan Kemp, London: Lawrence & Wishart, 1937.