Three Philosophical Problems about Consciousness and their Possible Resolution

Open Journal of Philosophy, 2011, volume 1, issue 1, pages 1-10.

Nicholas Maxwell
Emeritus Reader in Philosophy of Science at University College London
Email: *nicholas.maxwell@ucl.ac.uk*
Website: *www.nick-maxwell.demon.co.uk*

Abstract
Three big philosophical problems about consciousness are: Why does it exist? How do we explain and understand it? How can we explain brain-consciousness correlations? If functionalism were true, all three problems would be solved. But it is false, and that means all three problems remain unsolved (in that there is no other obvious candidate for a solution). Here, it is argued that the first problem cannot have a solution; this is inherent in the nature of explanation. The second problem is solved by recognizing that (a) there is an explanation as to why science cannot explain consciousness, and (b) consciousness can be explained by a different kind of explanation, empathic or "personalistic" explanation, compatible with, but not reducible to, scientific explanation. The third problem is solved by exploiting David Chalmers' "principle of structural coherence", and involves postulating that sensations experienced by us – visual, auditory, tactile, and so on – amount to minute scattered regions in a vast, multi-dimensional "space" of all possible sensations, which vary smoothly, and in a linear way, throughout the space. There is also the space of all possible sentient brain processes. There is just one, unique one-one mapping between these two spaces that preserves continuity and linearity. It is this which provides the explanation as to why brain processes and sensations are correlated as they are. I consider objections to this unique-matching theory, and consider how the theory might be empirically confirmed.

Keywords: Consciousness, mind-body problem, brain processes, explaining consciousness, functionalism, experiential functionalism, physical explanation, empathic understanding, sensation-brain correlations, unique matching theory, hard problem of consciousness.

1 The Three Problems

I am inclined to think that there are three basic philosophical[1] problems that arise in connection with consciousness.

(1) The Problem of Existence. Why does sentience or consciousness exist at all? Why are we not zombies?[2]

(2) The Problem of Intelligibility. Granted that consciousness exists, what is it? How is it to be explained and understood? On the face of it, there could be no greater mystery than that brains should somehow produce, or *be*, our states of awareness, our thoughts, feelings, perceptions and desires. What is so baffling and mysterious about consciousness is that each one of us knows it exists, and knows what it is, because we possess it, indeed we *are* it, in a certain sense; and yet, if we examine a conscious brain, we find such things as neurons and synaptic junctions, but nothing remotely like consciousness as we experience it. Consciousness is wholly apparent to the owner of the conscious brain, but bafflingly invisible and ineffable to everyone else. How is this familiar and utterly inexplicable stuff of consciousness to be explained and understood?

(3) The Problem of Explaining Brain-Mind Correlations. What possible explanation could there be for the way brain processes and sensations are correlated?

In what follows I suggest solutions to two of these problems, and indicate why, in my view, the other problem has no solution, and thus does not deserve to be regarded as a legitimate problem.

2 Functionalism, if Correct, Solves the Three Fundamental Problems of Consciousness

If functionalism is correct, all three problems are solved at a stroke. According to functionalism – as I think it ought to be formulated – the mental aspect of brain processes is simply what may be called the "control" aspect, that aspect involved in guiding the animal or person to act in the way that they do.[3] Viewed from a Darwinian perspective, the function of the brain is to control the animal to act in ways conducive to survival and reproductive success in the given environment. In referring to sensations, perceptions, feelings, desires, states of awareness, imaginings, thoughts, decisions to act, we are referring to neurological processes going on in the brain from the standpoint of their role in guiding or controlling action: detecting bodily changes or aspects of the environment (sensation and perception), assessing significance and prompting appropriate kind of response (feeling), determining or influencing choice of goals (desire), registering the current environmental situation (awareness), or exploring possibilities (imagining); and so on. According to functionalism, the mental aspect of brain processes is nothing more than this kind of control aspect.

This means that any brain, of whatever constitution or structure, that is sufficiently sophisticated to produce action just like the actions of a conscious person, thereby has a conscious, mental aspect just like the conscious, mental aspect of *our* brains, the brains of conscious persons. A zombie who behaves like a conscious person *is* a conscious person. Philosophical zombies do not, and cannot, exist.

One slight qualification to this conclusion can be recognized by functionalism. It is just about conceivable that a robot without a brain transmits radio signals to a vast, very rapid computer, which calculates what the robot would do, on the basis of received information, were it to have such and such a brain, and then transmits instructions to the robot as to how it should act. The robot acts as if conscious, as if it had a conscious brain, but in this case no such brain exists, but only a model of it in the computer, and so the robot is not conscious. It is a zombie.

For functionalism, then, there is no philosophical or conceptual problem concerning the *existence* of sentience or consciousness – or rather, in so far as there is a problem, functionalism solves it.

Functionalism also solves the problem of intelligibility, the problem of understanding what the nature of consciousness is. Sentience and consciousness are no more than the relevant control aspects of brains sufficiently sophisticated to produce action that we would describe as "sentient" and "conscious" in character.

And functionalism also solves the third problem, the problem of what possible explanation there can be for the way brain processes and sensations are correlated. The mental aspect of a brain process is given by the role that process plays in guiding the animal or person to act in the way he or she does (possibly taking counterfactual situations into account). Correlations are between the neurological processes, described as neurological processes, and these processes described in terms of the control role they play in producing actual and potential actions. What the control role of a neurological process is will depend on such things as its physical or neurological character, how it is situated in the brain, what the overall functioning structure of the brain is, what other functionally described brain processes the given process can, in part, cause to occur, when the rest of the brain is in this or that state. There is, in short, according to functionalism, no big mystery, no philosophical or conceptual problem, about why the neurological and mental aspects of brain processes are correlated in the ways that they are. There are, of course, immense and highly intractable empirical problems about *how* precisely neurological and mental (or control) aspects are correlated, made all the more difficult to solve by the complexity of the conscious brain, and by the moral objections to investigating the conscious brain in an invasive manner. Functionalism highlights the importance and

intractability of these empirical problems,[4] but disposes of the problem of how there could possibly be an explanation for brain-mind correlations. Given functionalism, there is no such problem.

Thus functionalism, if correct, disposes of the three fundamental philosophical problems of consciousness at a stroke. No wonder it is a popular view.

3 Functionalism is Not Correct

Functionalism is, however, untenable. A simple, well known argument shows decisively that functionalism cannot be correct. The argument goes like this.

Functionalism is put forward as a part of the reductionist programme of natural science, and can legitimately be assessed in that light. What functionalism achieves, if correct, is to show that there is nothing associated with conscious brains which lies irredeemably beyond the scope of scientific explanation. The mental aspect of brain processes is no more than the control aspect which will, one day, be explained and understood in neurological terms, in terms of brain structure and functioning, which in turn will be explained and understood in biological, chemical and molecular terms and, ultimately, in principle (if not in practice), in physical terms. We are physical systems put together by evolution to function in extraordinary ways, but nevertheless in ways that are ultimately, in principle at least, fully explicable physically. The brain is just another organ, with its specific function, like the heart, the lungs, the stomach or the liver: one day science will give us just as good an explanation of the structure and functional aspect of the brain as it does at present of the other organs.[5]

But physics, and that part of natural science in principle reducible to physics, cannot conceivably predict and explain fully the mental, or experiential, aspect of brain processes. Being blind from birth – or being deprived of ever having oneself experienced visual sensations – cannot in itself prevent one from understanding any part of physics. It cannot prevent one from understanding the physics of colour, light, physiology of colour perception and discrimination, just as well as any normally sighted person. In order to understand physical concepts, such as mass, force, wavelength, energy, spin, charge, it is not necessary to have had the experience of any particular kind of sensation, such as the visual sensation of colour. All predictions of physics must also have this feature. In order to understand what it is for a poppy to be red, however, it *is* necessary to have experienced a special kind of sensation at some time in one's life, namely the visual sensation of redness. A person blind from birth, who has never experienced any visual sensation, cannot know what redness is, where redness is the perceptual property, what we (normally sighted) see and experience, and not some physical correlate of this, light of such and wavelengths, or the molecular structure of the surface of an object which causes it to absorb and reflect light of such and such wavelengths. It follows that no set of physical statements, however comprehensive, can predict that a poppy is red, or that a person has the visual experience of redness. Associated with neurological processes going on in our brains, there are mental or experiential features which lie irredeemably beyond the scope of physical description and explanation. Functionalism is thus shown to be false.[6]

I might mention in passing that this argument, usually attributed to (Nagel, 1974) and (Jackson, 1982, 1986), was actually first put forward by me several years before Thomas Nagel and Frank Jackson, in two papers published in 1966 and 1968.[7]

There are two other arguments, in addition to the above colour-blindness argument, regularly employed by philosophers to establish the incompleteness of physics, the falsity of functionalism. There is, first, the inverted spectrum argument: it is conceivable that a person might make the same colour discriminations as I do, and might have a similar physiology, but might experience an inverted spectrum, seeing red when I see blue, and blue when I see red.[8] In the two cases, the physics is essentially the same, but the experience is different; hence physics cannot be complete. Second there is the zombie argument: it is conceivable that I have a twin, precisely the same as me physically, but devoid of consciousness: hence physics is incomplete.[9]

The three arguments are related to one another: the colour-blindness argument considers

sensorially deprived persons; the inverted spectrum argument considers the possibility of people with different sensory experiences; and the zombie argument considers the possibility of complete sensory deprivation. The great advantage of the colour-blindness argument over the other two, however, is that it alone does not rely merely on what is conceivable or possible: we know there are people who are colour-blind; we can consider what happens when people who are blind from birth have their sight restored. These are actualities, not mere possibilities.

The three arguments have been much discussed. Some hold onto functionalism and reject the arguments;[10] others hold that the arguments are valid, and reject functionalism.[11] Here I assume that one or more argument is valid, and functionalism has been shown to be false.[12]

At once we are confronted again by the three philosophical problems of consciousness with which we began (given that there is no other obvious candidate for the solution). In what follows I sketch a two-aspect theory of consciousness which, I claim, can be deployed to solve the second of the above three problems, the problem of intelligibility. I then put forward a unique-matching theory which is able to solve the third of the above three problems, the problem of explaining brain-mind correlations.

4 Experiential Functionalism

Before us there is, let us suppose, another conscious or sentient being, whether person, animal, alien, or even, possibly, robot or android. What is this utterly mysterious sentience or consciousness, associated with the brain processes of the other being? Why does it resist scientific explanation? How is it to be explained and understood?

Sentience or consciousness, according to the two-aspect view I wish to defend, is that aspect or feature of a brain process that we can only get to know about as a result of having a sufficiently similar brain process occur in our own brain. It is what it is to have that kind of process occur in one's own brain. It is just that, and nothing more.

This thesis, note, does justice to the baffling privacy of consciousness, expressed above in problem (2). If the mental aspect of a brain process is just what we get to know about in having that process occur in our own brain, then of course we cannot discern the mental aspect if the process occurs in another's brain. In order to discern the mental aspect it is necessary and sufficient to ensure that a sufficiently similar process occurs in our brain (assuming our brain is sufficiently similar to the other brain). However hard we peer at another person's brain, and however probing and thorough our investigation, we will never, in that way, detect the faintest hint of sentience or consciousness.

But if I want to know what the other being is experiencing in having a brain process, N, occur in his brain, how "sufficiently similar" a brain process, M, must occur in my brain (and how "sufficiently similar" must my brain be)? There are at least six possibilities.

(i) N and M are precisely the same physically, even if the two brains are not precisely the same.

(ii) N and M are precisely the same neurologically (i.e. the same pattern of neurons fire in the same way), even though there are otherwise differences between the physical states of the neurons.

(iii) Neurons may be quite different physically (e.g. in one case neurons are biological, in the other case made out of microchips), but the pattern of firing of the neurons, and the interconnections between the neurons, is the same.

(iv) "Strength of signal" may be coded in quite different ways at the neuronal level (so that in one case this is related to rapidity of firing of neurons, while in the other case it is related to strength of electric current, let us suppose); once these differences are ignored, however, the pattern of signals is the same in the two cases.

(v) The *functional* or *control* role of the neurological processes, N and M, are identical in the two brains, even though the pattern of signals, the "code" at the neuronal level, and the physical structure and functioning of the neurons, are entirely different.

(vi) The behaviour of the two beings is similar, even though the control architecture of the two brains is entirely different so that, from a *functional* or *control* standpoint, the neurological processes,

N and M, work in quite different ways.

(i) and (ii) require such a high level of similarity between N and M that they probably imply that we never ourselves have the same kind of experience on different occasions. (vi) requires such a low level of similarity between N and M that it is indistinguishable from behaviourism. The robot, considered above, that has no brain but is controlled by a computer to act as if it is conscious, satisfies (vi); but even functionalism, let alone the two-aspect view being considered here, can give reasons for holding the robot is not conscious. We are left with (iii), (iv) and (v). It is not easy to see how, even in principle, we could obtain evidence to decide between these options. Here, without argument, I plump for option (v). Sensations are to be correlated with the *control aspect* of brain processes – brain processes *functionally described*. This version of the two-aspect view might be called "experiential functionalism".[13]

But if mental features, correlated with brain processes described in control or functional terms, really do exist, why do such mental features lie beyond the scope of physics?

Physics does not, even in principle, predict and explain such a mental feature because physics is concerned only with those features of things that need to be referred to in order to predict how states of affairs evolve with the passage of time. Physics, in other words, is concerned exclusively with what may be called the "causally efficacious" aspect of things: see (Maxwell, 1968a; 1998, 141-55). Features of things which do not need to be referred to in order to predict future states of physical systems, will not be referred to by physics.

Suppose that the world is such that there is a yet-to-be-discovered, unified, explanatory, true physical "theory of everything". Suppose further, to keep the argument simple, that this theory is deterministic and classical in character. Such a theory is comprehensive and complete – a theory of *everything* – because, given any isolated system, the theory, together with a precise specification of the instantaneous physical state of the system (formulated in the highly specialized, restricted vocabulary of the theory), predicts future states of the system, described in terms of the same causally efficacious (i.e. physical) properties. But to say this is not to say that the theory predicts everything about the system, all facts about the system. If the system includes a conscious being, the comprehensive physical description of the system will include a precise specification of the physical state of the being's brain. But in order to carry out the predictive task of physics there will be no need to refer to the mental aspect of the brain, what it is to have that kind of process occur in one's own brain. As a result, the "theory-of-everything" will make no mention of such a mental feature.

But could not the physical "theory-of-everything" be extended so that it includes reference to mental features, and thus becomes a genuine theory of everything? Let it be conceded that this can be done. The crucial point to appreciate is that the new, amplified theory would be so horribly complex and ad hoc that it would entirely cease to be explanatory. Given the vast richness and complexity of the experiential world, and given the mind-boggling complexity of the manner in which even the most elementary of mental features, such as the visual sensation of redness, are correlated with physical states of affairs, the unity and explanatory power of the physical "theory-of-everything" would be entirely lost in the amplified theory. The amplified theory would consist of millions, possibly billions, of distinct postulates linking physical and mental features, each postulate itself of incredible complexity. All this would be in striking contrast to the fundamental simplicity, unity and explanatory power of the physical "theory-of-everything".

There is, in short, an *explanation* as to why physics does not, and cannot, include the mental, the experiential. If it did, the extraordinary explanatory power of physical theory would vanish. Excluding the experiential is the price we pay for having the marvellously explanatory theories that we do have in physics.[14]

It should be noted that the silence of physics about the experiential provides no grounds whatsoever for holding that the experiential does not exist. It is only if we hold that the only properties that exist are causally efficacious properties (or aspects of such properties) that this conclusion follows. But the thesis that only causally efficacious (or physical) features exist seems

decisively refuted by our experience of colours, sounds, smells and other sensational features – features which can only be known as a result of oneself having certain kinds of sensations, and hence certain kinds of (functionally described) brain processes occur in one's own brain.

It is rather natural to suppose that the stubborn resistance to scientific explanation exhibited by sentience and consciousness is due to, and is a sign of, their inherent mysteriousness and inexplicability. This is, indeed, a major reason for supposing that the mental is inherently inexplicable: the mental is so mysterious that it resists above all our very best kind of explanation, namely scientific explanation. But the supposition is wrong. Sentience and consciousness (like perceptual properties out there in the world, such as colours and sounds) evade scientific explanation, not because of any stubborn inexplicability, but because they are not required for the kind of causal explanation that science provides, and cannot be incorporated into science because, if they are, science (or at least that part in principle reducible to physics) ceases to be explanatory.[15]

But if sentience and consciousness cannot be explained and understood scientifically, how are they to be understood? Elsewhere I have argued at some length that mental features can be explained and understood by what may be termed "empathic" or "personalistic" explanation and understanding, a kind of explanation different from, compatible with, but not reducible to, scientific explanation.[16]

To understand another empathically or personalistically is to know what it would be like to *be* the other person (or sentient being), experiencing, feeling, thinking, believing, desiring, planning and deciding what that other person experiences, feels, etc. This involves arranging to occur in one's own brain neurological processes that are sufficiently similar (in control terms) to the processes that are occurring in the other person's brain, without this leading to one actually *doing* what the other is doing. We imagine that we are the other being. Imagination, quite generally, is arranging to occur in one's own brain processes sufficiently similar to those that would occur were one actually doing what one imagines one is doing.

Personalistic explanation is fundamentally anthropomorphic in character, and thus fundamentally distinct from scientific explanation, which is not anthropomorphic. Scientific understanding never involves relating what is to be understood to oneself, in an essential way. If, however, I want to understand another conscious being, an alien let us suppose, *as a person*, then it is essential that I bring myself into the picture, and relate the other to myself. If I want to know what the other, the alien, is experiencing, perceiving, feeling, etc., what I want to know is what it would be like for me to *be* the other, having processes occur in my brain that are sufficiently similar, in the relevant respects, to what occurs in the alien's brain. If I want to understand what the alien says or thinks, then I must discover how to *translate* the alien's language into mine. In every case, mental or personalistic features of the alien, which lie beyond the scope of science, are known, and can only be known, by bringing oneself into the picture and relating the other, the alien, to oneself – by understanding the other anthropomorphically, in other words.[17]

According to the psycho-functionalist version of the two-aspect view, indicated above, the mental features of brain processes are precisely the kind of features to be explained and understood personalistically. The mental feature of a brain process is what we know about in having a sufficiently similar process occur in our own brain; personalistic explanation and understanding is a kind of understanding quite specifically designed to enable us to understand just such a feature.

My solution to the second philosophical problem of consciousness, then, comes in two parts. First, a major reason why consciousness seems inherently inexplicable is because it seems to be inherently beyond the scope of even our best kind of explanation, namely scientific explanation. There is, however, an *explanation* for the incapacity of physics, and science reducible to physics, to explain consciousness: physics is concerned only with the causally efficacious aspect of things, and if physical theory is amplified to include the experiential, it might, in principle, be predictive, but it would cease entirely to be explanatory – for reasons that can be entirely explained and understood. Consciousness resists scientific explanation, not because it is inherently mysterious and inexplicable, but because it is the kind of thing which science can ignore, given its predictive task, and must

ignore, if it is to be explanatory. Second, consciousness, the mental aspect of brain processes, *can* be explained and understood, namely by means of personalistic explanation, a kind of explanation that is compatible with, but not reducible to, scientific explanation. The mental aspect of that kind of brain process that is the visual sensation of redness cannot be understood scientifically, but it is wholly understandable personalistically, for those of us with normal colour vision: we understand what it is, personalistically, in having that kind of process occur in our own brain.

This argument requires that personalistic explanation, even though not reducible to scientific explanation, is an intellectually authentic mode of explanation in its own right. Elsewhere[18] I have put forward arguments in support of this thesis. I have argued that it is the evolution of our human capacity for personalistic understanding that has transformed mere sentience into consciousness, and made our human language and culture possible (construed in personalistic terms). Even science presumes personalistic understanding in that scientists, in order to understand each other's ideas, problems and theories, must see things imaginatively from each other's perspective, thus acquiring a kind of etiolated personalistic understanding of each other (one that emphasizes beliefs about aspects of the environment and ignores most of the personal dimension). Science is thus, in a sense, based on personalistic understanding. Personalistic understanding is not folk psychology – construed to be a pre-scientific psychology which will be replaced by a genuinely scientific psychology as knowledge advances.

What may be called "teleological" or "purposive" explanation constitutes a *third* kind of explanation. This explains by interpreting the actions, or growth, of something as being designed to achieve goals. It is a watered-down version of personalistic explanation, in that actions are explained as being designed to realize aims, but no attempt is made to enable one to know what it would be like to *be* the thing in question: all reference to the mental or experiential is thus excluded (unless construed in purely functional terms). It applies equally to thermostats, robots, guided missiles, and all living things. In my view, everything can in principle be understood physically; all living things, and all purposive things created by us can, in addition, be understood purposively; and sentient things are open, in addition, to being understood personalistically. If functionalism were correct, then all that we are, as persons, could be explained and understood purposively, and personalistic understanding would be redundant. It is the incompleteness of functionalism, its failure to encompass fully the mental, which requires that purposive explanation be enriched so that it becomes empathic or personalistic explanation. (Experiential functionalism enriches, and does not just reject functionalism as traditionally conceived.)

5 Explaining Correlations and the Unique-Matching Theory

So much for my solution to the second of the three philosophical problems about consciousness, with which we began. But what about the third problem? How could it be possible to explain correlations between brain processes and sensations? Given existing correlations which we may presume hold between brain processes and visual sensations of colour, why should not the spectrum of colour sensations be reversed, so that the sensation of redness is now correlated with that brain process that was correlated with the sensation of blueness, and so on?

We have seen, in effect, that no amplification of scientific theory could explain why correlations between brain processes and sensation are as they are; any physical theory-of-everything amplified to include the experiential will be just as non-explanatory whether one considers colour sensations as they are, or as they would be given a reversed colour spectrum.

Furthermore, it seems that no amplification of personalistic explanation could do the trick either. Suppose there is a God-like brain, that can experience all possible sensations; suppose further that He knows all there is to know about the way His brain processes and sensations are correlated. Despite this vast store of experience and knowledge, it would seem that the God-like being is in no better position to explain why brain processes and sensations are correlated as they are than we are.

Until fairly recently, I found this argument convincing. And then it struck me that there is just one

circumstance in which an explanation for brain-sensation correlations *does* exist.

Suppose that a God-like brain is indeed possible; it is able to experience all possible sensations. Suppose, further, that our visual, auditory, tactile, olfactory and other sensations form disjoint regions within the continent of sensations experienced by the God-like being. To us, the different modes of experience seem utterly distinct in kind. Visual sensations seem utterly different from auditory ones, which again seem utterly different from tactile and olfactory sensations. If one had only experienced visual sensations, one could never guess what auditory, olfactory or tactile sensations would be like. But let us suppose that the God-like being is able to experience endlessly many sorts of sensations that lie between our visual, auditory, olfactory and tactile sensations. To Him, as he moves from the visual towards the auditory, there is a continuous, slight change in the quality of the sensations experienced, much as there is for us, within the auditory, when a tone goes continuously up in pitch. Moving, by means of continuous changes in the quality of sensations, from visual to auditory, the auditory comes as no surprise: it emerges as a result of smooth transitions, as in the case of the tone rising in pitch. And the same goes for transitions between other modes of sensations: the visual, the auditory, the olfactory, the tactile and so on. For the God-like being, all possible sensations lie in a multi-dimensional "space" of possible sensations, the experienced quality of sensations varying smoothly, continuously, indeed in a steady, linear way, as one moves around in the multi-dimensional "space" (one kind of smooth variation in sensation, such as changes in pitch of a sound of specific timbre and loudness, corresponding to one dimension in the "space" of sensations). To the God-like being, this vast realm of possible sensations has a kind of overall coherence, a unity, an overall structure, based on the fact that the quality of sensations varies smoothly, from sensation to sensation. There is, we may suppose, just one way in which all these possible sensations can be ordered and "placed" in the multidimensional space, so that continuity is preserved throughout the space, so that the closer together any two sensations are in this space then the more nearly alike they are experientially.

And let us suppose, further, in accordance with David Chalmers' "principle of structural coherence",[19] that all this is mirrored in the "space" of the functionally described brain processes that *are* the sensations. As the God-like being moves smoothly from experiencing one sensation to experiencing a slightly different sensation in a neighbouring "place" in the multidimensional space of all possible sensations, so the brain process, that is the first sensation, becomes the slightly different brain process that is the second sensation, slightly different, that is, when described in functional or control terms. The smooth coherence and unity of sensation space is matched by a corresponding smooth coherence and unity of brain process space, when brain processes are characterized in control terms. There is, as mathematicians would say, something like an isomorphism (a common structure) between the space of sensations and the space of brain processes. This matching of structure is, we are to suppose, unique. It can only be done in one way. Change the way sensations are matched up with (functionally described) brain processes, and the common structure between the realm of sensations and the realm of brain processes is lost. Even a rigid "rotation" of the two realms, the two spaces, with respect to each other, cannot be done.[20]

If all this is the case, as a possibility, then an explanation *is* possible as to why sensations and brain processes are correlated in the way that they are. They have to be correlated in this way, because if they are not, the matching of structure, of unity, based on continuity, between sensations and brain processes, would be lost.

If this proposal holds up, then the third philosophical problem of consciousness has been solved.

6 Objections

I now consider objections to the above unique-matching explanation for brain-sensation correlations, considered as a possibility.

A first reaction might be that the theory is much too wild and crazy to be a serious candidate for truth (even possible truth). I am almost inclined to agree with this assessment. One should note,

however, that again and again in the history of thought, ideas that initially seemed wild and crazy have subsequently become solid, almost prosaic items of knowledge.

On similar lines, the idea may be held to be far too vague and speculative to be taken seriously as a potential contribution to knowledge. But, again, ideas that were initially vague and speculative, have subsequently become important contributions to knowledge: atomism is an example. Elsewhere (Maxwell, 1998, especially 7 and 80-89) I have argued that "blueprints" – vague ideas for future theories – play a vital role in physics. In view of its vague, open-ended character, the unique-matching idea ought perhaps to be called a blueprint rather than a theory.

It may be objected that the above unique-matching explanatory theory (or blueprint) is extremely limited in scope, in that it applies only to sensations shorn of all intellectual content, and does not apply to thoughts, to feelings, to perceptions, imbued with intellectual content of one kind or another.

But this objection misses the point. What is hardest of all to explain and understand is why sensations that are stripped of intellectual content, "raw feels" as they have been called, are correlated in the way they are with brain processes. This is the fundamental mystery. When it comes to thoughts, feelings and perceptions imbued with intellectual content one *expects* correlations to exist between head processes described "mentalistically" or personalistically, on the one hand, and described in functional or control terms, on the other hand. An explanatory theory, of the kind indicated above, would be all the better for targeting the hardest part of the problem.

It may be objected, again, that the above matching- structure explanatory theory, even if in some sense correct, would nevertheless be largely incomprehensible to us human beings since we, unlike the purely notional God-like being, have not experienced, and cannot experience, all the sensations that intervene between the isolated patches of sensation, visual, auditory and so on, that we do experience. The theory "explains" at the price of being incomprehensible.

I have five remarks to make in response to this objection.

First, a theory of sentience or consciousness, couched in purely functional or control terms, omitting all reference to the experiential or personalistic, would predict that the notional God-like being (if it existed) would find it increasingly difficult to discriminate between sensations (functionally described), as these become closer together in the "space" of all possible (functionally described) sensations. (I am assuming, here, that the unique-matching theory is correct.) The functional theory of sentience would tell us, in effect, that for the God-like being, sensations are continuously arrayed in the "space" of all possible sensations, but would not tell us *what* the God-like being would experience, where we have not ourselves had the corresponding sensations.[21] Such a functional theory of consciousness would, of course, be entirely understandable to us.

Second, the matching structure-theory, if correct, would apply to the isolated patches of sensation that we *do* experience, visual, auditory, etc. We can now discern structure in these distinct modes of sensation, and the unique-matching theory would predict the existence of this structure, and match it up with corresponding structure in the distinct spaces of possible brain processes of a kind that occur in (normally experiencing) human brains. We would be able fully to understand this part of the unique-matching theory, applicable to us, even if we would not be able to have personalistic understanding of that part which applies to the God-like being but not to us.

Third, it is inherent in the very idea of an explanatory theory that it should predict new phenomena. The more it does this, so the more powerfully explanatory it is, other things being equal. In the case of the unique-matching theory, predicting new phenomena means predicting the potential existence of sensations of which we have had no experience, and probably cannot experience. That the theory does predict the potential existence of these "new" sensations, smoothly joining up visual, auditory, olfactory sensations and so on, is, it should be noted, essential to its explanatory capacity. Thus this feature of the theory, far from being a defect, is intrinsic to the explanatory power of the theory, an inevitable and even desirable feature.

Fourth, it deserves to be noted that when one views the matter from a Darwinian perspective, it

would seem not implausible that there should be a continent of potential sensations smoothly varying from visual to auditory to olfactory, etc. It is important to survival that a sentient animal does not confuse vision with hearing, with smell or with touch; natural selection would favour changes in brain structure that make distinct modes of sensation as *distinct* as possible, and different sensations within a mode as distinct as possible. It is not unreasonable to suppose, in other words, that our brains have been designed by Darwinian evolution to deliver to us distinct patches of sensation as *different* from one another as possible. Darwinian theory makes the matching structure-theory seem not unreasonable. Fifth, the solution to the philosophical problem of how it is possible to explain correlations between brain processes and sensations being offered here does not require that our isolated, distinct modes of sensation really are connected up by smoothly varying, but unknown sensations in a vast space of possible sensations: all that is required is that this is a *possibility*, not that brains really could exist that would experience all these sensations unknown to us. But it is, of course, always possible that the unique-matching theory, or something like it, really is the *correct* explanation for mind-brain correlations.

It may be objected, yet again, that the above matching- structure explanation does not explain, and cannot explain, why *this* particular brain process correlates with (or *is*) *this* particular sensation, the visual sensation of greenness, let us say, or the sound of a flute softly playing middle C. But this is just what the theory *would* explain. Such and such a kind of brain process (described in functional or control terms) cannot be other than the sound of a flute softly playing middle C because, if it is anything else, the isomorphism between the space of (linearly varying) brain processes, and the space of (linearly varying) sensations, is lost. Matching of smooth structure requires that this particular brain process be correlated with this particular sensation, and not some other sensation.

It may be objected that the unique-matching theory, if true, must be true analytically; it cannot, therefore, be explanatory. Suppose the theory is true to the extent that the continuously varying space of all possible sentient brain processes exists. The closer together two points in this space are, the more difficult the God-like being finds it to discriminate between the corresponding sensations. This much a correct, comprehensive, but purely functional theory of the brain will be able to predict, without any appeal being made to the sentient or phenomenal. But if this is the case, then it must follow, analytically, that as points become close in brain process space, so corresponding points become close in sensation space. One simply cannot have dramatically different sensations, and yet find it very difficult to discriminate between them. So if the space of all sentient brain processes (functionally described) is continuous, then it follows, analytically (all but logically) that the same must be true of the matching, correlated space of all possible sensations.

But what this shows is that the unique-matching theory, if true, is powerfully explanatory; it does not establish that the theory itself is analytic. The theory postulates that all (functionally described) brain processes that are sentient can be arrayed in a space which is such that brain processes vary smoothly, linearly, everywhere, in the manner indicated. This is a factual postulate which may well be false. The unique-matching theory is thus also factual, and not true analytically; it may well be false. But *if* the space of all possible sentient brain processes varying linearly everywhere exists, then the corresponding space of all possible sensations must be phenomenologically linearly varying everywhere as well and, to this extent, there must be an explanation as to why brain processes and sensations are correlated in the way that they are.

It may be objected that even if the unique-matching theory is true of the God-like being that experiences all possible sensations, it would still be possible for us human beings to have brain processes and sensations correlated in a way that is different from the way they are actually correlated; indeed, each one of us may, for all we know, have quite different sensations even though we have roughly similar brains and make the same discriminations. None of this would refute the unique-matching theory, interpreted as being about the God-like brain; hence the theory cannot explain brain-sensation correlations in us.

My reply is that a basic presupposition of the unique-matching theory is that experiential

functionalism is correct. The mental or experiential aspect of a brain process occurring in another's brain is what it is like to have that kind of process (described in functional terms) occur in your own brain (granted that your brain is sufficiently similar functionally to the other's brain). The presupposition is, in other words, that functionally sufficiently similar brains, that are conscious, experience the same kind of sensations. This is a presupposition of the unique-matching theory, and hence something that the theory cannot explain. What the theory *can* explain (if true) is why all functionally sufficiently similar conscious brains all have the particular sensations that they do have, and not sensations that are, one and all, locally different (so that all experience colours reversed, for example).

It is important to appreciate that there are limitations to what the unique-matching theory can explain. It cannot explain why functionally sufficiently similar conscious brains all have the same kind of sensations, as opposed to quite different sensations. Nor can it explain why the God-like brain experiences the sensations He does, and not sensations that are all systematically different. And nor can the theory explain why there are sensations at all, and not just zombie-like functional or control aspects of brains.

Another objection that may be made is that the God-like brain, able to experience all sensations, may well be quite impossible, even in principle. It is important to appreciate that, given a more or less specific, localized kind of brain process, this may well have quite different control and experiential features associated with it, depending on how it connects up with the rest of the brain, and depending on what the character of the rest of the brain is. Given this, there is a problem as to how the God-like brain could possibly have sensations similar to ours – our brains being very different from, and much smaller than, the God-like brain. It may be, in other words, that it is quite impossible for a single brain to exist that can experience all possible sensations. But if this is the case, what can it mean to say of the space of all possible sensations that, in it, sensations vary continuously as one moves through the space? Must we not, rather, think of the space of all possible sensations as being made up of isolated islands rather than a single continent, each island corresponding to a particular kind of conscious brain, comparisons of sensations from different islands being meaningless?

Even if the isolated island hypothesis is true, the unique-matching theory could still have an explanatory role to play. It could be that the "isolated island" that encompasses all sensations experienced by human beings is such that, sensations vary continuously as one moves through the "isolated island" space. A brain is possible that can experience all these sensations. In this case, the unique-matching theory would explain why we cannot experience sensations different from the ones we do experience, taken from our "isolated island" space of sensations. The theory would not explain why we cannot have sensations taken from another "isolated island" space – all the sensations of the two islands being swapped. It could be, however, that some other explanation can be found as to why sensations cannot be swapped wholesale between "isolated islands".

There is another possibility. Even if the space of all possible sensations does split up into islands, each island corresponding to a specific kind of functionally described possible sentient brain, it could still be that these islands overlap. Overlapping would make it meaningful to speak of sensations varying continuously throughout the space of all possible sensations. In this case, even though the God-like brain is not possible, the unique-matching theory nevertheless applies to all possible sensations.

Another objection, made in a helpful spirit by David Chalmers during an email conversation, is that symmetries that obtain in the space of human sensations would have their counterparts in the space of all possible sensations, and hence the unique-matching theory cannot be correct. An example of a symmetry in our sensation space is the inverted colour spectrum: colour sensations are inverted but colour discriminations remain the same.

My reply to this objection comes in three parts.

First, the space of all possible sensations, postulated by the unique-matching theory, does have a

kind of symmetry – a kind of position invariance: wherever you are, sensations vary linearly in all directions. But this is of course different from the kind of discrete symmetry that Chalmers has in mind, involving something like a local reflection or rotation. Ignoring the space of all human sensations for the moment, are there grounds for holding that the space of all possible sensations must possess discrete symmetries? What prevents rigid rotations of the whole space, for example? This could be prevented by the space having a jagged enclosing boundary, excluding the possibility of rigid rotation. If the jagged enclosing boundary is modified to form an enclosing surface of a hypersphere, which would make rigid rotations possible, then sensations no longer vary at the same rate throughout the space in all directions. But suppose the boundary is fixed, and suppose points in the space are moved so as to preserve continuity: what would prevent that? The answer would have to be that, after the continuity-preserving transition, the rate of change of brain processes with changes of position would no longer keep pace, throughout, with the rate of change of sensations with position. Linearity would be violated.

Second, Chalmers' objection can be interpreted to be that it is not logically possible for the space of human sensations to exhibit discrete symmetries (such as colour reversal), and for such discrete symmetries to be absent in the space of all possible sensations. But this is surely wrong. Postulate that the space of all possible sensations exists, sensations changing in a continuous and steady (i.e. linear) way with changes in position. Postulate that there are no local or global discrete symmetries. Such a space seems logically possible. Given this space, it is clearly going to be possible to select out widely separated sub-spaces to constitute all possible human sensations. Furthermore, this can be done in such a way that, given this collection of sub-spaces, one can easily imagine the possibility of discrete symmetries, such as colour reversal, obtaining, without this destroying continuity (linearity) in the space of all possible sensations. No contradiction seems to result from postulating a continuous space of all possible sensations, in which there are no discrete symmetries, and a collection of sub-spaces, which forms all possible human sensations, within which discrete symmetries are possible.

Third, Chalmers' objection may be interpreted to be, more modestly, that it is implausible to suppose that the space of all possible sensations does not exhibit discrete symmetries, especially in view of the fact that the space of all human sensations does exhibit such symmetries. My reply to this is that at present we know next to nothing about the character of the space of all possible sensations, and certainly not enough to be able to declare that the hypothesis of continuity, or linearity, is implausible. Furthermore, if we accepted that the space of all possible sensations is globally linear, then it is entirely to be expected that the space of human sensations will *not* have this character, and will in fact exhibit discrete symmetries. As I have already pointed out, it is vital for survival that sensations are such that different sensations can be quickly and easily distinguished, both within a sensory mode, and between modes. Fruit eating creatures need to be able to distinguish red from green easily, so that ripe fruit may be readily seen among green leaves. It would be disastrous if an animal got confused as to whether sensations were visual, auditory or tactile. It is entirely to be expected, in other words, that natural selection will select out sub-spaces from the space of all possible sensations, to form a creature's sensations, that are widely separated in the continuous, or rather linear, space of all possible sensations. To form an animal's sensations from some small, homogeneous region in the space of all possible sensations would be unhelpful, even disastrous, from the standpoint of survival. Darwinism, in other words, ensures that, IF the linear space of all possible sensations exists, THEN the space of all possible sensations of this or that creature that is a product of evolution, will not be linear or continuous, and will exhibit discrete symmetries. That some kind space of all possible sensations exists, larger than the space of all possible human sensations, larger even than the space of all possible sensations of sentient creatures on earth, seems, on the face of it, entirely reasonable. It is the further requirement that this space is linear throughout that must be regarded as uncertain and speculative.

Finally, it may be asked: what evidence can be gained in support of the conjecture that the space

of all possible sensations exists, with the crucial feature of linearity required for the matching structure-theory to be true? It is difficult to see how we can explore this conjecture *experientially*, without submitting ourselves to thoroughly objectionable and radical brain surgery. It is possible, however, that we may one day develop a functional theory of sentience and consciousness, which meets with empirical success as far as human brains are concerned (and perhaps also other mammalian brains), and which makes predictions about the character of the space of all possible sensations. Even hard-nosed functionalists ought, perhaps, to be interested in the speculative idea put forward here, because it has implications for even a purely functionalist theory of consciousness.

This concludes my discussion of objections to the unique-matching explanation as to why brain processes and sensations are correlated as they are. If this theory is at least possibly true, then this solves the third of the above three philosophical problems of consciousness, with which we began.

7 Can the Existence of Consciousness be Explained?

So far I have suggested solutions to the second and third philosophical problems of consciousness, but hardly anything has been said about the first problem, the problem of explaining the existence of consciousness.

My claim is that this problem has no solution, and is thus not really a problem at all. Any explanation can only explain X by showing that it is a manifestation of something else, Y, in terms of which X can be explained. In asking for an explanation as to why anything whatsoever exists, rather than nothing, one deprives oneself of anything, any Y, in terms of which the explanation is to be couched: this problem is a non-problem. In the physical realm, we cannot reasonably expect to be able to explain why the physical universe exists, rather than nothing, because this provides us with nothing, no Y, in terms of which the existence of the universe may be explained. (This remains true even if a version of inflationary big bang cosmology is true, which asserts that the cosmos emerged from the vacuum as a runaway quantum fluctuation. The physical vacuum, or pre-vacuum is not *nothing*.) So, too, in the experiential realm, we cannot explain why the experiential exists, rather than there being nothing experiential. In this case, we cannot appeal to the physical, and explain the experiential as emerging from the physical for, as we have seen, the experiential cannot be derived from the physical. We can, perhaps, explain the way the physical and the experiential are correlated; but it is inherent in the very concept of explanation that neither the existence of the realm of the physical, nor the existence of the realm of the experiential, is capable of being explained. In both cases, the problem is a non-problem.

8 Conclusions

Here are a few lessons which, in my view, can be drawn from the above discussion.

One big division in the community of those who seek to understand consciousness is between those who hold, roughly, that functionalism will one day make sense of consciousness, and those who hold that functionalism is false or incomplete, there being what David Chalmers and others have called "the hard problem of consciousness", the problem of explaining consciousness in a sense which goes beyond functionalism. The view that emerges from the above discussion, differs somewhat from both these orthodox positions. Traditional functionalism is indeed radically incomplete. Nevertheless, understanding what functionalism leaves out is not quite as "hard" as it is sometimes taken to be. We do already have a mode of explanation – personalistic explanation – which does enable us to explain and understand features of consciousness, which we experience directly, and which functionalism leaves out. The really "hard" problems of consciousness, on this view, are the functionalist problems: the problems of linking up brain processes described in neurological terms, and in control or functionalist terms – terms which we can relate to our personalistic understanding of ourselves and each other. As these "hard" functionalist problems are solved, this will undoubtedly have major implications for personalistic understanding. The result will be the enrichment and improvement of personalistic understanding, not its replacement by something

better (as some proponents of folk psychology have claimed).

If the unique-matching theory, sketched above, cannot be shown to be untenable, then it deserves to be taken very seriously, as demonstrating that it *is* possible to explain brain-mind correlations.[22] If the theory turns out to be untenable then the way forward, in my view, is to take seriously the arguments designed to show that no scientific, indeed no, explanation of brain-mind correlations is possible, and search for counter-examples.[23] This is the strategy that I have somewhat blindly put into practice; it has led me to stumble across the unique-matching idea indicated here. Finally, in tackling the philosophical mind-body problem, and indeed all philosophical problems, we should try to come up with solutions which have fruitful consequences. We should keep before us the great example of Darwin, who succeeded in solving a profound philosophical problem – the problem of how purposive living things can have proliferated in a purposeless universe – in a way which has had endless fruitful implications for the whole of biology, and for our understanding of ourselves.[24]

References

Armstrong, D. M. (1968). *A materialist theory of the mind*. London: Routledge and Kegan Paul.

Block, N. (1990). Troubles with functionalism. In W. Lycan (Ed), *Mind and cognition* (pp. 444-468). Oxford: Blackwell.

Campbell, K. (1970). *Body and mind*. New York: Doubleday.

Chalmers, D.J. (1996). *The conscious mind*. Oxford: Oxford University Press.

Clark, A. (2000). *A theory of sentience*. Oxford: Oxford University Press.

Dennett, D. (1991). *Consciousness explained*. London: Allen Lane.

Jackson, F. (1982). Epiphenomenal qualia. *Philosophical Quarterly 32*, 127-136.

Jackson, F. (1986). What mary didn't know. *Journal of Philosophy 83*, 291-295.

Kirk, R. (1974). Zombies versus materialists. *Aristotelian Society 48 (Supplement)*, 135-152.

Lakatos, I. (1976). *Proofs and refutations*. Cambridge: Cambridge University Press.

Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly*, *64*, 354-361.

Lewis, D. (1972). Psychophysical and theoretical identifications. *Australasian Journal of Philosophy 50*, 249-258.

Lewis, D. (1990). What experience teaches. In W. Lycan (Ed.), *Mind and cognition* (pp. 499-519). Oxford: Blackwell.

Locke, J. (1961). *An essay concerning human understanding*. London: J. M. Dent & Sons.

Lockwood, M. (1989). *Mind, brain and the quantum*. Oxford: Blackwell.

Maxwell, N. (1966). Physics and common sense. *British Journal for the Philosophy of Science 16*. 295-311.

Maxwell, N. (1968a). Can there be necessary connections between successive events?. *British Journal for the Philosophy of Science 19*. 1-25. Reprinted in R. Swinburne (Ed.) (1974), *The justification of induction* (pp. 149-174). Oxford: Oxford University Press.

Maxwell, N. (1968b). Understanding sensations. *Australasian Journal of Philosophy 46*. 127-146.

Maxwell, N. (1984). *From knowledge to wisdom: A revolution in the aims and methods of science*. Oxford: Blackwell, (2nd ed., 2007, revised and extended. London: Pentire Press).

Maxwell, N. (1985). Methodological problems of neuroscience. In D. Rose and V.G. Dobson (Eds.), *Models of the visual cortex* (pp.11-21). Chichester: John Wiley.

Maxwell, N. (1998). *The comprehensibility of the universe: A new conception of science*. Oxford: Oxford University Press.

Maxwell, N. (2001). *The human world in the physical universe: consciousness, free will and evolution*. Lanham, Maryland: Rowman and Littlefield.

Maxwell, N. (2010). *Cutting god in half – and putting the pieces together again: A new approach to philosophy*. London: Pentire Press.

Mulhauser, G. (1998). *Mind out of matter*. Dordrecht: Kluwer.
Nagel, T. (1974). What is it like to be a bat?. *The Philosophical Review 83*. 435-450.
Nagel, T. (1986). *The view from nowhere*. Oxford: Oxford University Press.
Nemirow, L. (1990). Physicalism and the cognitive role of acquaintance. In W. Lycan (Ed.), *Mind and cognition* (pp. 490-499). Oxford: Blackwell.
Place, U. T. (1956). Is consciousness a brain process?. *British Journal of Psychology 46*, 44-50.
Popper, K. R. (1962). *The open society and its enemies, vol. II*. London: Routledge and Kegan Paul.
Putnam, H. (1960). Minds and machines. In S. Hook (Ed.), *Dimensions of mind* (pp. 138-164). London: Collier-Macmillan.
Rey, G. (1997). *Contemporary philosophy of mind*. Oxford: Blackwell.
Smart, J. J. C. (1963). *Philosophy and scientific realism*. London: Routledge and Kegan Paul.

Notes

1. By a philosophical problem I mean a conceptual problem so baffling that we don't know whether or not it is a serious problem of knowledge and understanding, there being no agreement as to what the problem is or what would count as a solution. One basic task for philosophy is to try to clarify serious philosophical problems so that they turn into fruitful empirical, or solvable, problems: see final section.
2. See (Campbell, 1970; Kirk, 1974).
3. Some functionalists take the analogy with the computer and the Turing Universal Machine very seriously, and hold that a major task is to give definitions to interconnected psychological terms: see (Putnam, 1960; Lewis, 1972; Rey, 1997). The latter is a characteristic obsession of analytic philosophy: for a corrective see (Popper, 1962, 9-21). I am inclined to see functionalism as a minor modification of the kind of brain process theory defended by (Place, 1956; Smart, 1963; & Armstrong, 1968) – a modification which stresses that the mind is to be identified with the functional or control aspect of the brain, the stuff of the brain being irrelevant. This viewpoint takes the basic task to be to solve the philosophical mind-brain problem, it being important to put this problem into the context of biology and evolution: see (Maxwell, 1984, 174-181 & 269-273); and (Maxwell, 1985). The argument of the present paper does not, however, require that functionalism be interpreted specifically in this "control", Darwinian fashion.
4. One of the great virtues of functionalism is that it transforms the apparently utterly inexplicable philosophical mind-body problem into a problem that is, in principle, an empirical problem, however intractable it may be. This virtue is retained by the modified "experiential functionalist" view I defend below.
5. Some have argued, however, that functionalist phenomena are not reducible to physics: see, for example, (Block, 1990, 446; Maxwell, 2001, ch. 6).
6. The nub of this argument is that there are real features in the world (sensory qualities as we experience them) which are such that, if F is such a feature, then it is necessary oneself to have experienced a specific kind of F-sensation if one is to know what F is, which means (according to the view to be defended below) that it is necessary to have had a specific kind of brain process, functionally specified, occur in one's own brain. No physical property is like this. Hence physics cannot predict F-type features, if they exist. On the face of it, they do exist: colours, sounds, smells as we experience them (whether interpreted as being without or within us) appear to be just such features. Some rebuttals of the argument, such as (Mulhauser's, 1998, ch. 4), fail to come to grips with the argument when formulated as above.
7. See (Maxwell, 1966, especially 303-308); and (Maxwell, 1968b, especially 127, 134-137 & 140-141). When I recently drew Thomas Nagel's attention to these publications, he remarked in a letter, with great generosity: "There is no justice. No, I was unaware of your papers, which made the central point before anyone else". Frank Jackson acknowledged, however, that he had read my 1968 paper.

8. This possibility was first discussed by (Locke, 1961, bk. 2, ch. 32, section 15). For discussion see (Dennett, 1991, 389-398; Chalmers, 1996, 99-101).
9. Versions of this argument have been discussed by (Kirk, 1974) and (Campbell, 1970): the argument is rejected by (Dennett, 1991), and endorsed by (Chalmers, 1996, 94-99).
10. (Nemirow, 1990; Lewis, 1990; Dennett, 1991, 398-406; Clark, 2000, 30-35).
11. (Maxwell, 1966; 1968b; 1984, 259-73; 2001, ch. 5; Nagel, 1974; 1986; Chalmers, 1996, chs. 3 & 4).
12. There is also the so-called "argument of the explanatory gap" – see (Levine, 1983) – designed to show that physics cannot explain the experiential. But it seems to me that the main reason for believing in the "explanatory gap" is the "colour-blindness" argument, sketched in the text.
13. Versions of this view have been defended by (Maxwell, 1966; 1968b; 1984, 174-81 & ch. 10; & 2001) and (Nagel, 1974; 1986). For a particularly lucid and detailed exposition of the view see (Chalmers, 1996, especially chs. 6 and 7).
14. See (Maxwell, 2001, ch. 5) for a fuller exposition of this argument.
15. An anonymous referee has raised two questions about this argument that there is an explanation as to why science cannot explain consciousness. First, the very question "Why does consciousness exist?" is ambiguous. It could mean either "What is the cause of consciousness?" or "What is consciousness for?". I have argued that there is an explanation as to why there cannot be an explanation for the existence of consciousness when the question is interpreted in the first way. As for the second interpretation, it is clear physics, and that part of science in principle reducible to physics, cannot answer "What is consciousness for?", because physics does not deal with function and purpose. But evolutionary biology, a part of science, does. Perhaps Darwinism can supply an explanation when the question is interpreted in this fashion? Orthodox Darwinism seems incapable, however, of providing an answer. For any such answer would, it seems, have to appeal to the survival value, or reproductive fecundity, of consciousness, and any such explanation could not, it seems, explain why zombie consciousness would not serve these ends just as well as real consciousness. I have however, elsewhere, distinguished nine different versions of Darwinism, which progressively give greater and greater roles to purposiveness, sentience and consciousness, to the theory: see (Maxwell, 2010, ch. 8). The last three of these appeal to sentience, and to what may be called "sentient explanations", and the last two to consciousness and "personalistic" explanations, but this hardly amounts to a Darwinian explanation for the evolutionary emergence of sentience and consciousness, as the context makes clear. These explanations appeal to sentience and consciousness when they may be presumed to have evolved, but do not explain why they evolved in the first place (a non-zombie version, that is). Secondly, the referee raises the question of whether physics can even explain *biological* phenomena, let alone phenomena associated with sentience and consciousness. There is, however, in my view, a fundamental difference between the incapacity of physics to explain biological phenomena, conceived of in zombie terms, and phenomena associated with consciousness. Nothing going on in connection with biology, conceived of in zombie terms, cannot in principle (we are entitled to assume) be explained physically. It is just that physics cannot supply us with certain *kinds* of explanation – purposive and personalistic – as I go on to make clear. But the situation seems to be very different when we come to sentience and consciousness. Here there seem to be features of brain processes, our inner experiences, which seem to be utterly beyond the scope of physics, and of science, unless some part of science – psychology, for example – is interpreted in a special way from the outset so as to include inner experiences and consciousness.
16. See (Maxwell, 1984, especially 181-189 & 264-275) and (Maxwell, 2001, chs. 5-9). See also (Maxwell, 2010, chs. 3, 7, and especially 8).
17. See also (Maxwell, 2001, 114-115).

18. See note 16.
19. Chalmers' "principle of structural coherence" asserts that, as far as human brains and states of consciousness are concerned, structural features of brain process space match structural features of conscious experience space: see (Chalmers, 1996, 222-225). (Lockwood, 1989, 109-210) indicates a similar idea. The unique-matching theory is, in effect, a particular application of Chalmers' principle.
20. I first put this idea forward in (Maxwell, 2001, 126-129).
21. (Chalmers, 1996, 233-242) makes a related point in connection with his "principle of structural coherence".
22. Science quite generally, I have argued (Maxwell, 1984; 1998), must make metaphysical assumptions concerning comprehensibility or explainability. A "science" of consciousness must do likewise.
23. (Lakatos, 1976) has argued, brilliantly, that one use of a proof is to aid the search for counterexamples.
24. See (Maxwell, 2001, 168-179) for a development of this point.