CHALMERS ON THE ADDITION OF CONSCIOUSNESS TO THE PHYSICAL WORLD

1 Introduction

In The Conscious Mind, David Chalmers argues that the facts about the distribution and character of conscious experiences in the world are not logically fixed by the physical facts -- or as he puts it, consciousness does not logically supervene on the physical.[1] Furthermore, a number of passages suggest that he holds a stronger view, denying that the facts about conscious experiences are even logically constrained by the physical facts. For example, he lists solipsism, panpsychism, and the theory that people are conscious only in odd-numbered years among theories logically compatible with the data that an individual person has access to (216). The view he argues for is that consciousness naturally supervenes on the physical, where the relation between consciousness and the physical world is governed by fundamental bridge laws that he claims can be expected to be as simple as the fundamental physical laws. On the standard understanding of such natural supervenience, consciousness is a kind of addition to the physical world. Chalmers is also attracted to a version of the view in which primitive phenomenal or protophenomenal properties play a role in constituting the intrinsic nature of the physical.[2] I shall set aside this admittedly speculative position in this paper, though some of what I say may apply to it.

Part 2 of this paper begins with a preliminary section pointing out a number of difficulties for the view that consciousness naturally supervenes on the physical without being logically constrained by it. It is radically epiphenomenal, involves unexplained coincidences, and would lead to a very great increase in the complexity of our picture of the universe. Not only would it require a new category of fundamental fact, but the new facts would be extremely numerous and complex. Then in the main section I argue that some thought experiments taken or adapted from those Chalmers uses in fact show that the way that consciousness depends on the physical world is logically constrained to a considerable degree. In part 3 I argue against the weaker, logically constrained, natural supervenience by

suggesting that the most plausible account of these logical constraints is that consciousness is entirely logically fixed by the physical world, and by presenting modified versions of some of the difficulties for the not logically constrained natural supervenience view. Finally in part 4 I distinguish the two types of logical supervenience, which I call conceptual and metaphysical, arguing that metaphysical supervenience is no more plausible than natural supervenience, and sketching some lines of defence against some of Chalmers' objections to the conceptual supervenience view.

The formulation of logical supervenience that I wish to work with claims that mental properties are logically supervenient on physical properties if and only if for every mental property M and every x such that x is M, there is some physical property P such that x is P and a bridge law P-->M (saying that if something is P then it is M) holding in all logically possible worlds with our physical laws.[3] If M is a phenomenal property, then P may need to include only the intrinsic physical properties of things which have it. But if M is an intentional property, then to accommodate wide content, P must include relational as well as intrinsic properties of things which have it.[4] I shall call bridge laws in which the antecedent is a more fundamental kind of property 'supervenience laws'. (So, for example, if functional properties supervene on physical properties there will be supervenience laws of the form P-->F, where F is a functional property.)

I shall refer to such logical supervenience as conceptual supervenience when, given our physical laws, the supervenience laws P-->M hold as a matter of a priori, conceptual necessity. And I shall label separately as 'metaphysical supervenience', the position in which these supervenience laws hold as a matter of a posteriori necessity and not of conceptual necessity.[5] I depart from common usage here in taking conceptual supervenience not to entail metaphysical supervenience. And I depart slightly from Chalmers in taking 'logical' always to mean metaphysical or conceptual, whereas he uses 'logical' and 'conceptual' interchangeably (and argues that metaphyiscal supervenience requires conceptual supervenience.) Natural supervenience requires that such supervenience laws are contingent

a posteriori, holding with a weaker natural (or nomological) necessity. I depart from common usage again here in taking logical supervenience not to entail natural supervenience. This position takes the supervenience laws in effect to hold in all possible worlds with our natural laws -- the physical laws plus any further fundamental laws there might be, such as psychophysical supervenience laws (36).[6]

Few people dispute that mental properties supervene on physical properties. Where the disagreement lies is over the status of the supervenience laws -- whether they hold as a matter of conceptual, metaphysical, or natural necessity.[7]

A popular way of looking at the logical supervenience of a mental property M on physical properties, suggested by Saul Kripke, is to say that God would not have to do anything more to the universe to establish in it some instances of M than simply establish the physical facts.[8] But if M naturally supervenes on physical properties, God would have to establish instances of M by some further act, e.g. by decreeing that there be fundamental supervenience laws P-->M, or F-->M where F is a functional property. (It is clear that functional properties conceptually supervene on physical properties, since as functional states are defined in terms of causal relations among physical and other functional states, there can be no difference in functional properties without a difference in physical properties. So there will be conceptually necessary laws P-->F for all F, which will combine with F-->M laws to yield P-->M laws.) The difference between conceptual and metaphysical supervenience is that supervenience laws P-->M would be explainable if conceptual but involve essentially unexplainable components if metaphysical.

It now seems hard to dispute that special science properties such as those of chemistry, geology, meteorology and biology all conceptually supervene on physical properties in this way. We find nothing puzzling or implausible about the idea that if a certain complex physical property P obtains in a region, then an oxidation, volcanic eruption, thunderstorm or heart attack must be happening there as a matter of conceptual necessity. Once the physical antecedent is stated there are no options for what the consequent might be.

Even God would be powerless to create any. The concepts given by the antecedent and consequent together with the particular set of physical laws seem obviously to determine a supervenience law once the special science concept is seen as structural or functional. So these supervenience laws do not supplement the fundamental truths about our world. God would not have to do anything beyond establishing all the physical facts of the universe in order to bring about oxidations, volcanic eruptions, thunderstorms and heart attacks, .[9]

Let us turn now to the mental properties. Many hold that intentional mental properties also conceptually supervene on physical properties (at least in the case of those intentional properties that make no reference to phenomenal properties). This may be because they believe intentional properties conceptually supervene on functional properties. For, as mentioned, it is clear that functional properties conceptually supervene on physical properties.

But many have argued against the conceptual supervenience of phenomenal properties on physical properties or, as I shall sometimes say, the conceptual supervenience of consciousness on the physical. And their arguments have been conveniently collected together and forcefully presented by Chalmers. Central among these is the idea that in the case of P->M laws in which M is a phenomenal property we find it all too easy to imagine P being instantiated but not M. For we may sometimes be able to imagine some other phenomenal property, and we can always imagine no phenomenal property at all. Such imaginability is then taken to imply conceptual possibility. Suppose, for example, that M is the property of being a sensation of greenness. It is then argued that any P-->M law cannot be conceptually necessary because it is imaginable and hence at least conceptually possible that P could be instantiated in worlds with our physical laws by something which is a sensation of redness, and that in these worlds people could accordingly have their phenomenal colour spectrum inverted relative to ours. Similarly it is argued that it is imaginable and hence conceptually possible that in a world physically identical to ours our twins have no sensations at all -- they would be what philosophers now call zombies.

This completes my initial presentation of the three candidates for the supervenience of consciousness on the physical, the general appeal of the position that all properties conceptually supervene on the physical, and the special difficulties conceptual supervenience faces with consciousness. I'd now like to set out what I take to be the most striking difficulties for the view that consciousness supervenes naturally upon the physical without being logically constrained by it -- the position adopted by Chalmers.

2.1 Difficulties for the Natural Supervenience of Consciousness without Logical Constraints

First, on the standard version of this view phenomenal properties are radically epiphenomenal in that there are logically possible worlds physically identical to the actual world but with all the consciousness removed. (This is a much stronger and seemingly less plausible epiphenomenalism than that frequently charged against nonreductive materialisms, regardless of the strength of their supervenience, on the grounds that they treat not mental states but their physical realisations as doing all the causal work.[10] For on those views there need be no logically possible worlds in which phenomenal states such as pains could be removed without removing their physical realisations.) Chalmers concedes that the epiphenomenalist nature of the standard version of his position is counterintuitive. But he takes it not to be a fatal flaw (160), and argues that epiphenomenalism can be avoided on the other version (153-5).

Second, if the distribution of consciousness in the world is not logically constrained by the physical but instead is governed by fundamental natural laws, then its actual location in the world amounts to an extraordinary unexplained coincidence. It is a coincidence, for example, that consciousness happens to be attached at the only place in which consciousness talk can be interpreted unproblematically as referring to consciousness. Chalmers formulates two plausible principles that he takes to serve as natural constraints on the fundamental laws of consciousness (in Chapter 6). These principles are (i) the coherence principle -- where there is consciousness, there is awareness, and where there is the right kind of awareness, there is consciousness, and (ii) the structural coherence principle -- structural features of

consciousness correspond directly to structural features that are represented in awareness. Awareness Chalmers defines as a state 'wherein we have access to some information, and can use that information in the control of behaviour (28)'.

Consider first the plain coherence principle. Cast into theological language, Chalmers' view would allow that God could have attached consciousness to nothing at all or to anything he wanted, but of all the possible options he chose to attach it to awareness. Actual theists might herald this as a new argument for design.[11] A benevolent and hence non-deceptive God would have good reason to make consciousness and awareness cohere. But if we cannot assume a benevolent motive in selecting where to attach consciousness, and its attachment to awareness is governed by or derivable from a fundamental psychophysical or psychofunctional law of the universe, then this is a remarkable coincidence. What we have is a fortuitous association between consciousness and awareness, similar to Leibniz's psychophysical parallelism, and similarly in need of a benevolent God to make it credible. Likewise with the structural coherence principle. According to Chalmers we would have chance or divine design to thank or blame for the fact that our visual experiences exhibit the structure they do when we visit an art gallery.

One might respond that there is no extraordinary coincidence to be explained because in any world in which consciousness is attached to X it will seem particularly apt that it is so attached, purely because of familiarity. But it is questionable that it will seem that consciousness is attached to X at all, let alone that the attachment to X will seem familiar, in worlds physically and functionally identical to ours in which consciousness is attached to some X other than awareness. For example, in the world in which consciousness is attached to awareness in odd-numbered years and nowhere in even-numbered years, this alternation will go unnoticed and so will not seem familiar or apt.

Third, if phenomenal properties naturally supervene on physical properties, then there will be facts about phenomenal states which are not logically determined by physical facts about the universe. The physical laws and initial physical conditions of the universe will

then need supplementing with fundamental laws governing the distribution and character of consciousness in the world. As these supplements would be neither boundary conditions nor laws of succession but bridge laws, they would introduce an entirely new species of fact into the catalogue of fundamental facts governing our universe. And if, as Chalmers claims, they would be needed only to account for consciousness, this further reduces the plausibility that such a fundamental category exists.

A fourth difficulty, which I want to discuss in greater depth, is that these supplementary laws will be individually complex and extremely numerous, leading to a greater complexity in our picture of the universe. (If phenomenal properties do not even naturally supervene on physical properties, e.g. if there is spectrum inversion within our world, the picture of the universe will be still more complex as it will require perhaps individual statements of the presence of phenomenal states whenever they occur.) Chalmers recognises that the fundamental laws governing consciousness will have to be a simple set if his view is to be plausible. But he thinks they may well come to be seen as such. We can expect, he says, that these laws, like the fundamental laws of physics, could be written on the front of a T-shirt (214).

So let's consider how complex this set of fundamental laws will have to be. We need not assume that every phenomenal property must appear in the consequent of one of these laws. For example, we might exclude the property of being the taste of a nectarine, believing that it is analysable into the taste of a peach and the taste of a plum. Instead I shall begin by assuming that the consequents of these fundamental laws will be a set of fundamental phenomenal properties in terms of which all phenomenal properties can be derived. How many such fundamental properties will be needed to cover the entire phenomenal space of the universe? Perhaps a place to begin is to consider whether different sensory modalities can be covered by the same fundamental phenomenal properties. It is a commonplace that sensations of redness are not at all like sensations of the sound of a trumpet, and this reflection suggests we will at least need different fundamental phenomenal properties for

different sensory modalities. There are at least five human sensory modalities. And we can expect to keep counting to cover the variety of sensory modalities existing in terrestrial and extraterrestrial creatures, actual and possible.

Now consider the number of fundamental properties that will be required for each sensory modality. Beginning with human taste, we know that the great variety of what are commonly thought of as discernible tastes are really influenced by the sense of smell. But setting aside the influence of smell, it was thought until recently that all human tastes could be reduced to four -- salty, sour, bitter, and sweet. If this were right, then these four properties would be all that are needed to represent the phenomenal space of taste. At any rate it is hard to see how some more frugal and primitive set of phenomenal properties would allow us to derive these four cardinal tastes. So it appears that there will be at least four fundamental properties needed for human taste. Now think of the different kinds of smells that humans can recognise (and we know that many animals are much more discriminating). It is unlikely that they can be systematised into a small class of primary smells in terms of which all smells can be analysed. And moving on to more complex sense modalities, it looks as though for humans alone there will be an enormous number of fundamental phenomenal properties. The nineteenth century psychologist Edward Titchener counted 44,345 fundamental sensations using a method of introspection.[12] Although we might be able to quibble with this figure, we can expect that the number would be much higher if it is to include the phenomenal properties of nonhuman animals, extraterrestrials, and possible creatures.

So if the fundamental <u>laws</u> governing consciousness have fundamental phenomenal properties as consequents, there will need to be at least as many fundamental laws as there are fundamental phenomenal properties, and this will be a large number. But will there need to be any <u>more</u> than one fundamental law for each fundamental phenomenal property? To answer this we need to turn our attention to the antecedents of these fundamental laws. If these antecedents are to be couched in fundamental physical vocabulary, then according to

the widely accepted view that phenomenal states are multiply physically realisable, it would be very unlikely that there would be only one such physical antecedent P, for a given M, that could appear in a fundamental psychophysical supervenience law P-->M. Rather, it appears that if the fundamental laws governing consciousness have antecedents drawn from fundamental physical vocabulary, there will be many different fundamental psychophysical laws for every fundamental phenomenal property.

Chalmers doesn't think we are in a position yet to say much about how the fundamental laws governing consciousness will look. But he says they will be constrained by a principle of organizational invariance holding that if there is a functional isomorph of a conscious system, then it will have the same sort of conscious experiences (249). This tells us at least that the fundamental laws of consciousness needn't have fundamental physical antecedents, but need only have functional (or higher-level) antecedents. So the fundamental laws of consciousness might take the form F-->M, where F is a functional property. And assuming there are a multiplicity of P realising a given F, a multiplicity of nonfundamental P-->M laws will be derivable. But making the fundamental laws of consciousness psychofunctional laws would not be a whole lot better from the standpoint of simplicity. For it is increasingly held that functionalism faces multiple realisation problems of its own.[13] There appear to be a number of choices of inputs and outputs with which to construct functional properties, and even when such a choice is fixed there appear to be a multiplicity of functional properties that are sufficient for a given phenomenal property. It is unlikely, for example that a single functional property suffices both for your having a sweet-tasting sensation and for mine, because the having of this sensation can be expected to be connected, in a way that a functional characterisation must capture, to a cluster of beliefs not all of which we share. So it appears that there would be many more than one psychofunctional fundamental law needed for each fundamental phenomenal property.

A problem with fundamental laws of this form is that they fail to reveal what physical or functional properties determine nonfundamental phenomenal properties, and thus they fail

to cover all of phenomenal space. Rather than having a mapping onto the sensation of sweetness, for example, we need a mapping onto the sensation of sweetness of degree x, where x is a numerical variable. The degree of sweetness will be presented in the law as a function of variables occurring in the antecedent. If these variables are such as occur in physical and functional properties (e.g. voltage of an electric current, loudness of a scream), we can expect there to be a great many of them in each such law. Perhaps there will be fewer variables needed if instead of physical or functional properties we have some higher-level properties supervening on functional properties or, as Chalmers suggests, informational properties. But this is highly speculative. It should also be noted that although we can be confident that there are individual laws of the kind considered earlier for each degree of sweetness, there is no guaranteeing that there will be a single law using a variable ranging over degrees of sweetness that can replace such a system of laws.

We can now present the set of laws using numerical variables as a single law mapping a multi-dimensional nonphenomenal space into a multi-dimensional phenomenal space. And this may allow for some simplification if, for example, the function of variables yielding degree of sweetness is the same as, or related to, the function of variables yielding degree of loudness. However, even given the greatest conceivable simplification -- if each phenomenal parameter were to be presented as the same function of some set of nonphenomenal parameters -- we would still need to know for each fundamental phenomenal parameter, which set of nonphenomenal parameters it is a function of. This would still suggest a very implausible addition in complexity in our picture of the universe. For there would still be an enormous number of fundamental phenomenal parameters, and hence effectively an enormous number of individual laws. And the antecedents of these laws, if physical or functional, would very likely contain a great many variables. The complexity is both ontological and explanatory because it concerns both the contents of the universe -- fundamental laws and properties -- and what is unexplainable about it. This view certainly fails the T-shirt test.

One response Chalmers might make is to say that all phenomenal properties may be derivable not from what I've called fundamental phenomenal properties, but from some new kind of property that's neither mental nor physical, which he calls protophenomenal. The fundamental laws would then need only to connect physical and protophenomenal properties. And there might then be just a few such protophenomenal properties and hence just a few fundamental laws governing consciousness. This he concedes is pure speculation as we have no inkling what such a protophenomenal vocabulary would look like (127) and thus how the phenomenal vocabulary would be derivable from it. I think it is clear that any view that does not depend on accepting these yet to be articulated protophenomenal properties has a great advantage over a view which does require them.

2. 2 Some Logical Constraints on Consciousness

Having pointed out these problems for the view that consciousness is not logically constrained by the physical, I now wish to offer some positive demonstrations of logical constraints.

It may be hard to show that an inverted colour spectrum is logically impossible, but an extreme analogue of colour inversion can be shown logically impossible. Consider a world physically and functionally indistinguishable from the actual world in which the sensations of pleasure and pain are switched, so that the phenomenal feel of pain is associated with functional states of attraction, and the phenomenal feel of pleasure is associated with functional states of avoidance. In that world torture would be ecstasy and orgasm would be excruciating. Is such a world logically possible?

What are we to suppose happens in the pleasure/pain inverted world when our twins there produce marks and sounds 'I am in pain' and 'the quality of the experience I am having now is appropriate for my activity of avoiding such an experience'? Ex hypothesi such creatures are conscious. And if we understand introspection to be the process by which reflection upon one's phenomenal states issues directly in beliefs about them, these creatures are genuinely introspecting. Yet they do not notice that their orgasms are excruciating. For

it is unlikely we could find a plausible interpretation of their speech which would indicate otherwise. For example, if we took their 'pain' to refer to pleasure, and construed their 'appropriate' as meaning inappropriate, it is doubtful we could preserve the plausibility of our interpretation of their other utterances. And in any case, we would expect their noticing this misfit to show up in their nonverbal behaviour as well. Yet they resolutely stick to their ways, choosing orgasm over torture. Any unexpressed thoughts that there is something strange about their phenomenal experience would be detectable as brain activity, and ex hypothesi they are identical to us in this respect. So they do not notice the inappropriateness of the quality of their pains and pleasures.

Now I take it that we are certain that we do not inhabit such a world, because on at least some occasions in which we are in pain we are certain that the sensation we are experiencing is not a pleasant one. And I take it that this certainty is a special feature of pleasure/pain inversion that does not apply to colour spectrum inversion. We recognise that the phenomenal quality of pain is appropriate for flinching and avoidance behaviour in a way that is not just a matter of familiarity. It is unlike the case in which we familiarly associate the phenomenal quality of a sensation of redness with looking at ripe tomatoes. So because on some occasions in which we are in pain we can recognise that our sensations are appropriate for avoidance behaviour, we are certain that we are in pain.

But rational reflection upon the failure of our pleasure/pain inverted twins to notice the inappropriateness of their pains should lead us to question our claims to certainty. On occasions in which we are certain we are in pain, we must be certain we are not in the pleasure/pain inverted world. In order to have this certainty we must either be certain that the pleasure/pain inverted world is logically impossible, or be able to point to some difference between us and our twins that we are certain of that would justify our certainty that we are not in the pleasure/pain inverted world. What could this be? It cannot be that we are directly acquainted with our experiences, for ex hypothesi they are too. Nor can it be that we believe we are directly acquainted with pain, for ex hypothesi the same is true of them.

Nor can it be any combination of our beliefs for they will have the same beliefs as we do. Nor can we say it is the fact that we are directly acquainted with pain while they are directly acquainted with pleasure, for it is the certainty of our beliefs that we are in pain that is at issue. And for this reason also it could not be a combination of any of our beliefs together with the fact that we are directly acquainted with pain.

In the absence of such a difference between us and our twins that would justify our certainty that we are not in the pleasure/pain inverted world, the only way we can be justified in this certainty is by being certain that the pleasure/pain world is logically impossible.

Does a similar argument undermine the logical possibility of the zombie world -- by which I shall mean the world physically and functionally identical to ours but devoid of any consciousness? Our certainty from introspection that we are conscious would be undermined by the reflection that we are certain of no significant difference between us and the zombies which could justify our certainty that we are not in the zombie world. But there are differences between the zombie and pleasure/pain inverted worlds that might be exploited to block the argument in the zombie case. It might be argued that beliefs that one is conscious are incorrigible while beliefs that one is in pain are not. And it might be argued that while our pleasure/pain inverted twins have beliefs that they are in pain, our zombie twins lack beliefs that they are conscious, because there is no consciousness in their world.

We might try working our way to the logical impossibility of the zombie world by considering successive worlds with more and more of the phenomenal character removed from them. Consider first a world indistinguishable from ours except that with regard to pleasure and pain all experiences have a phenomenal quality we would describe as neutral. Such a world would be ruled logically impossible for the same reasons as the pleasure/pain inverted world. Now imagine a world in which, in addition, visual fields are always a blank grey, yet when people reflect on their grey visual experiences, they report them as having a colourful and complex content and make extensive use of what they report in their actions. This too would be ruled out. And by progressive steps one could eventually rule out the

world physically indistinguishable from ours but in which all experiences have a bland neutral quality. Call this the bland world. But it seems we can get no closer to eliminating the zombie world by removing any more phenomenal character. We can imagine partially zombie worlds in which one of the sensory modalities is zombified, e.g. one in which instead of visually experiencing a grey field people have no visual experiences at all. But then they couldn't be said to be introspecting when they report seeing their way through a busy street, for although they are conscious they are not having any visual experiences. So partially zombie worlds cannot be ruled logically impossible by the same arguments that ruled out the pleasure/pain inverted world.

For a further logical constraint, let us consider Chalmers' own scenario of the dancing qualia world, which he describes as 'so extreme that it seems only just logically possible (269)'. In this world, people's phenomenal states alternate back and forth between phenomenologically distinguishable states, e.g. between a sensation of redness and a sensation of greenness, perhaps every few seconds, without any change in physical and hence functional state. In such a world, someone looking at a field of grass is actually having a sensation of greenness for one second then a sensation of redness for the next second and so on alternating back and forth. Because this world is physically and functionally indistinguishable from ours, the person gazing at the field does not behave any differently as a result of seeing these alternating colours. He utters such words as 'That's a fresh green', just as I do. Now if this is logically possible, then the question arises how you and I can be sure that we are observing steady colours rather than alternating colours. As we are uncertain of any significant distinction between ourselves and our twins in the dancing qualia world, it seems that we must conclude that we cannot be certain that we are not in the dancing qualia world. But I think it is essential to our understanding of having a colour sensation, and experience in general, that we can be certain of whether a major change in our experience has just occurred. I think we are certain from introspection and intellectual reflection about various features of the quality of our experience not just at a mathematical instant but also for

a period of time. This is because even the quickest thoughts take time -- long enough for changes in one's experience to be apparent. While the duration of the specious present is undoubtedly a feature of the content of our thoughts, I take it also to be an essential feature of the thoughts themselves. Descartes' evil demon could not create a world in which there is a thought but no passing time.

Chalmers himself says that 'if we are to suppose that dancing qualia are naturally possible, we are led to a worrying thought: they might be actual and happening to us all the time (269).' While he uses this worrying thought to establish the natural impossibility of dancing qualia, I am arguing that it should lead us to reject the logical possibility of dancing qualia. To the idea that reflection on dancing qualia might be used to establish logical impossibilities Chalmers offers two responses (274). First, he says that it is disputable 'that it is constitutive of qualia that we can notice differences in them'. I have effectively been arguing that we should take it to be a constitutive feature of qualia that we can notice some changes in them. I shall be mentioning his second response later.

One might now seem to see a way of challenging the logical possibility of zombie worlds by constructing a variant of the dancing qualia world in which the phenomenal quality flickers back and forth with the absence of any phenomenal quality. In such a world people alternate between conscious and zombie phases. This flickering qualia world can be ruled logically impossible using the same kind of argument that ruled against the logical possibility of dancing qualia: We are not certain of any difference between ourselves and our twins in the flickering qualia world that would ground our certainty that our states of consciousness have endured through time. Yet our current specious present is long enough for us to be certain that we have been conscious during that time interval. We couldn't be certain of this if we weren't certain either that we can distinguish ourselves from our twins in the flickering qualia world or that the flickering qualia world is logically impossible. So the flickering qualia world is logically impossible. And we can play with the duration of the periods in which consciousness is switched on in this thought experiment, taking as an extreme example

the world in which I am conscious for a short period of time and the rest of the world is completely zombified. This world is logically impossible and perilously close to the zombie world. But again the zombie world itself escapes dismissal by this kind of argument.

To summarise, we may distinguish a strong denial of the logical supervenience of consciousness on the physical from a weaker denial. The strong denial is the position taken by Chalmers that the distribution of physical and hence functional properties imposes no logical constraints upon the distribution of phenomenal properties in the world. Once God has created the physical facts he is at total liberty to paint in phenomenal states if and wherever he wishes. On such a position, all I could be certain of from introspection is that I am conscious at a mathematical instant in time. I could not be certain from introspection that my stsate of consciousnes endures through time, or of anything about the phenomenal quality of my current conscious state. Nor could I be certain of the structure of my current conscious state, for I could not be certain I do not inhabit the bland world. There is a sense in which this would not be taking consciousness seriously. I find it an immensely counterintuitive position.

3 Arguments for the Logical Supervenience of Consciousness

The weaker denial of logical supervenience asserts that the distribution of phenomenal properties is logically constrained but not logically fixed by the distribution of the physical and hence functional properties in the world. This position would also count as natural supervenience since the fixing of phenomenal properties, within the constraints imposed by logical factors, would be natural. God would have more to do after establishing the physical facts of the universe in order to fix the phenomenal facts, but his hands would be at least somewhat tied -- he would be limited in his choices of where to paint in phenomenal states, and of the quality of those phenomenal states in given circumstances.

How plausible is this view that there are logical constraints on consciousness but limited ones which do not extend so far as to yield the logical supervenience of consciousness? Let us begin by considering whether any further logical truths might be

thought to follow from the logical impossibility of the pleasure/pain inverted world and the dancing qualia world.

From the logical impossibility of the pleasure/pain inverted world it follows that there could not be a logically possible world in which someone is in the very same physical state I am in when I am in pain but who is experiencing pleasure. From this it can be inferred that for those physical states I have been in on occasions in which I have been in pain, anyone in that physical state will experience an unpleasant sensation if she experiences any sensation at all. The qualification 'if she experiences anything at all' is needed because we have not ruled out the possibility of zombies in the same kind of physical state I am in who experience nothing at all. Let us make the further plausible assumption that a mere physical difference that makes no fine-grained functional difference cannot ground a difference in what a person is certain of that she might use to distinguish herself from her pleasure/pain inverted twin. Then returning to the theoretical standpoint of this paper, in which some kind of supervenience is being assumed, we can now assert

(A) For any instance in which someone is in pain, there is a fine-grained functional state F such that she is in F, and it is logically necessary that anyone in F experiences a phenomenal state within a certain range (e.g. unpleasant) if she experiences any phenomenal state at all.

For if it were not logically necessary, a person in pain introspecting could not be certain whether she was not actually experiencing the greatest of pleasures.

(A) rules out the logical possibility of the pleasure/pain inverted world. But so would (A') in which generalisation is made to include all phenomenal states:

(A') For any instance in which someone is in a phenomenal state M, there is a fine-grained functional state F such that she is in F, and it is logically necessary that anyone in F experiences a phenomenal state within a certain range of M if she experiences any phenomenal state at all.

The thought experiment obviously endorses (A') in the case where M is pain or pleasure. But it seems reasonable to extend it to other phenomenal states. For example, we might agree that an analogous thought experiment shows that a world in which visual sensations are

replaced by a space of painful sensations is ruled out, so that for visual sensations, the range of phenomenal states fixed by F in (A') at least excludes painful ones. In effect (A') derives from the thought that for every kind of phenomenal state we can be in, there are limits to the degree we can be mistaken about its qualitative character.

Now consider a stronger condition that would also rule out pleasure/pain inversion, in which generalisation is made to include all possible experiencers, and the range of phenomenal states is replaced by a single phenomenal state:

(A'') For any instance in which something is in a phenomenal state M, there is a fine-grained functional state F such that it is in F, and it is logically necessary that anything in F is in M if it is experiencing any phenomenal state at all.

(A'') is simpler and less arbitrary than (A') and in this respect is more plausible as a logical truth. But in both departures from (A') it clearly goes beyond what is revealed by the thought experiment. It applies to primitive experiencers who are unable to introspect (if there are any). And it claims that it's logically impossible to be in the same functional state that one is currently in, yet in a different phenomenal state. This would rule out spectrum inversion, and we do not yet have a thought experiment to accomplish this. However, the logical possibility of spectrum inversion, and hence the falsity of (A''), cannot be established by mere reflection. I shall return to this when discussing conceptual necessity.

A stronger claim still can be obtained by dropping the 'if it is experiencing any phenomenal state at all' from (A''), yielding the thesis of logical supervenience of consciousness on the functional:

(LSC) For any instance in which something is in a phenomenal state M, there is a fine-grained functional state F such that it is in F, and it is logically necessary that anything in F is in M.

(LSC) would be favoured over (A'') on grounds of simplicity, and unlike (A'') it would even rule out zombie worlds.[14]

Turning now to the logical impossibility of the dancing qualia world, a similar argument shows that

(B) For any instance in which someone is in M, there is a fine-grained functional state F such that she is in F, and it is logically necessary that if anyone in F is in phenomenal state M at a certain time, then for a time shortly beforehand during which she is continuously in F, she is in a phenomenal state not very different from M.

For if it weren't logically necessary, one couldn't be certain when experiencing a phenomenal state that it hadn't just changed greatly.

(B) explains the logical impossibility of dancing qualia. But so too would stronger conditions which entail (B). Generalise (B) to apply to the person at all times at which she is in F, replace 'not very different from M' by 'M', and we get (B'):

(B') For any instance in which someone is in M, there is a fine-grained functional state F such that she is in F, and it is logically necessary that if anyone in F is in phenomenal state M at a certain time, then at all other times at which she is in F, she is in M.

Both (B) and (B') would rule out dancing qualia without ruling out spectrum inversion within a world. (B') goes beyond what is revealed by the thought experiment, but is to be favoured on grounds of nonarbitrariness.

Now generalise (B') in two different ways to apply to all experiencers and we get (B''):

(B'') For any instance in which something is in M, there is a fine-grained functional state F such that it is in F, and it is logically necessary that if anything in F is in phenomenal state M at a certain time, then anything in F at any time is in M.

(B'') effectively says that the laws governing a world must consistently assign the same phenomenal state to a given functional state, but a different pairing of functional and phenomenal states is possible for other worlds. This would rule out worlds containing zombie and nonzombie twins and spectrum inversion within a world, but not the zombie world or spectrum inversion across worlds. Again this goes well beyond what is revealed by the original thought experiment, but is favoured on grounds of nonarbitrariness. Earlier I mentioned that Chalmers has a second response to objections to the logical possibility of dancing qualia. It is that even if dancing qualia were logically impossible, this might establish the logical necessity expressed in (B'') but it would establish no more.

The advocate of logical supervenience may now suggest taking (A') and (B) as the thin edge of the wedge and argue the steps to (LSC). Since each step involves a leap that is not supported by the original thought experiments and it is unlikely that another thought experiment will precisely fill this need, the best that can be offered is the following inference to the best explanation argument based on considerations of nonarbitrariness and simplicity.

If one sticks with (A') and (B), and does not endorse any of the stronger conditions, this looks implausibly arbitrary, as does endorsing any of the stronger conditions without endorsing the most general, (LSC). One cannot simply reject (LSC) on the grounds that it contains a logical necessity claim which is not obviously true, for the same could be said about (A') and (B). (LSC) would also be favoured as the only simple condition which rules out both dancing qualia and pleasure/pain inversion. Other thought experiments might reveal other highly specific kinds of logical constraints on consciousness, and considerations of simplicity would suggest that it is more plausible that there should be the single simple kind of logical constraint given in (LSC) than an unconnected patchwork of logical constraints.

This argument for logical supervenience is obviously not a conclusive one, and would be readily counteracted by opposing arguments. But before turning to these, let us consider whether we have a further argument against this logically constrained natural supervenience (and so <u>for</u> logical supervenience) by examining the extent to which the difficulties canvassed earlier for natural supervenience apply when we allow for logical constraints.

This depends on the extent of the logical constraints that have been identified. One difficulty for the unconstrained natural necessity view was the appearance of unexplained coincidences in its pairings of phenomenal states with physical states. To take care of this objection we would need to suppose that there are widespread logical constraints -- something like a logical construal of Chalmers' coherence and structural coherence principles. These principles are admittedly extremely nonspecific, but they would lend support to the more specific proposals for logical constraints that I have just been

considering. This would also remove the implausibility of epiphenomenalism. For although an epiphenomenalism would remain on this view, it would be restricted to a domain that might not be regarded as significant.[15] And the complexity problem would be diminished too. So far then, it appears that the more widespread the logical constraints, the better it looks for natural supervenience. But the implausibility of introducing a fundamental new category of fact just for consciousness is exacerbated if widespread logical constraints are acknowledged. For the role of fundamental psychophysical supervenience laws would no longer be the grand one of governing the presence of consciousness in the universe. Logical principles would already constrain the distribution of phenomenal properties in the world, just as they are widely regarded as wholly fixing the distribution of special science properties in the world. The fundamental supervenience laws governing consciousness would just be finishing off a job already started by logical principles--taking up the slack in fixing the distribution of phenomenal properties in the world. Like the God of the gaps, this natural supervenience of the gaps becomes increasingly implausible as its job is whittled away leaving its huge resources responsible for ever smaller and less significant aspects of reality.[16]

I have offered a number of arguments against the view that consciousness supervenes naturally on the physical. Still, these are not knock-down arguments. And they would have to be weighed alongside the arguments against the logical supervenience of consciousness. I cannot undertake this task here but will sketch a few thoughts, returning to the distinction between the two types of logical supervenience -- metaphysical and conceptual.

4.1 Problems for the Metaphysical Supervenience of Consciousness

Chalmers offers some arguments, with which I am largely sympathetic, for thinking that the supervenience of consciousness on the physical cannot be metaphysical. Rather than review them here, I shall offer a few supplementary points.

My approach will be to consider which of the difficulties discussed earlier for the natural supervenience view apply also to metaphysical supervenience. I shall begin by

comparing natural supervenience without logical constraints to metaphysical supervenience without conceptual constraints. As already explained, the special problem of epiphenomenalism facing the natural supervenience view does not apply in this case. But the remaining difficulties for the natural supervenience view do arise to some extent for this view.

First, it is acknowledged that there is a category of a posteriori metaphysical necessities made prominent by Kripke with examples such as 'if something is H_2O then it is water', where 'water' is understood as a rigidified designator. But if, as its adherents generally hold, the supervenience of consciousness on the physical does not derive from the use of a rigidified designator but is sui generis, then the type of metaphysical necessity that is required for the supervenience of consciousness indeed involves the introduction of a new category of truth. And as in the case of naturally necessary supervenience laws, there is something especially implausible about introducing a new fundamental category of truth which has application only in the case of consciousness.

Second, metaphysically necessary supervenience conditionals will still be individually complex and extremely numerous. As they are complex truths of a new category this can be construed as ontological complexity, though arguably the position needn't be regarded as introducing new fundamental properties. And the view is clearly explanatorily costly in the sense that there are in principle no explanations of these metaphysically necessary supervenience conditionals.

Third, the existence of unexplained coincidences is still a problem for this metaphysical supervenience view, since there would be no explanation in principle of the matching of consciousness to awareness. And this time God cannot even be brought in to rescue the position.

Fourth, this introduces a further problem unique to the metaphysical supervenience view. On this view God would be powerless to influence the way certain physical states are

felt, yet this would not be a limitation based on conceptual impossibility, but an entirely different and mysterious form of limitation.

Fifth, another new drawback of the metaphysical necessity view is that its advocates tend to argue for the supervenience of phenomenal properties on physical or functional properties by arguing for the identity of phenomenal properties with physical or functional properties, thereby picking up the burden of responding to the multiple realisability problem, or being prepared to embrace infinitely disjunctive properties.

And parallel to the case of natural supervenience with logical constraints, metaphysical supervenience is no more plausible if we allow conceptual constraints.[17] For again let us assume the conceptual constraints have been identified to the point at which they remove all appearance of unexplained coincidence from the pairings of phenomenal states with physical states. The complexity problem will be somewhat diminished but not significantly. The two problems unique to the metaphysical supervenience view will remain -- its mysteriousness and its invoking psychophysical or psychofunctional identities. And again, the implausibility of introducing a fundamental new category of fact just for consciousness is exacerbated if widespread conceptual constraints are acknowledged.

4.2 Defending the Conceptual Supervenience of Consciousness

On balance it seems to me that metaphysical supervenience is no better off than natural supervenience. That leaves conceptual supervenience to inherit the arguments for logical supervenience. Indeed, since those arguments depend on taking certain features of phenomenal states as constitutive, it appears that the logical impossibility of worlds such as the pleasure/pain inverted world discussed earlier is conceptual impossibility. Although this does not obviously extend to the logical necessities discussed in (A) etc, it is plausible to suppose that they too could be conceptual. Conceptual supervenience suffers from none of the difficulties raised against natural and metaphysical supervenience. For example, although on the conceptual supervenience view there will be a large number of complex

conceptual truths, they are neither ontologically nor explanatorily costly. But instead it faces the five formidable opposing arguments presented by Chalmers (94 -106).

I have already hinted at how one might respond to the objections from the apparent conceptual possibility of spectrum inversion and zombies. The strategy would be to argue that intuitions supporting conceptual possibilities are weaker than those supporting conceptual impossibilities. When something seems conceptually possible as far as one can tell, it is always open that one might find a thought experiment showing it after all not to be conceptually possible. But a single thought experiment might convince one of a conceptual impossibility. Thus one could plausibly maintain that spectrum inversion and zombies are conceptually impossible though not obviously so. The inference to the best explanation argument, though not a powerful one, nonetheless offers a reason for believing this.

The principal further arguments Chalmers offers against conceptual supervenience are the knowledge argument, including Frank Jackson's famous version of it, which I shall not discuss here, and an argument from lack of analysis, to which I now turn.

Chalmers argues that the lack of any kind of analysis of the notion of consciousness shows that a conceptual entailment cannot be forged between consciousness and the physical. The 'purported analyses do not even get into the ballpark. ... There is no temptation to even try to add epicycles to a purported functional analysis of consciousness in order to make it satisfactory... (106)'. It puzzles me why Chalmers suggests that an analysis, i.e. necessary and sufficient conditions, is required of consciousness in order to demonstrate the conceptual entailment from physical to phenomenal. For it is an attractive feature of his book that it focuses on questions of sufficient conditions of consciousness, setting aside the question whether such conditions are also necessary and the attendant problems of multiple realisability, heterogeneous disjunctions, and identity of mental and physical properties. It may well be that one could supply an account of the conceptual entailment from physical to phenomenal without being able to offer an analysis. All Chalmers needs to say is that we have no idea how the conceptual entailment could go. This is in fact how he begins his

section on the argument from the absence of analysis (104), claiming that those holding consciousness is entailed by physical facts are obliged to show how the entailment might possibly go.

In my view it is not essential to a defence of the conceptual entailment of consciousness by the physical to meet this challenge. One thing the consideration of mathematical examples such as Goldbach's Conjecture shows is that one can have good reason for holding that something is conceptually impossible without having a clue as to where the contradiction lies. In the Goldbach case, that good reason is that number crunching computers have tested the Conjecture for all natural numbers up to some very large number and have found no exceptions. It is entirely consistent with high confidence in the Conjecture that one have not the slightest clue how to prove it. Somewhat similarly, one might have good reasons for holding the conceptual supervenience of consciousness on the physical without having the slightest clue how phenomenal states are entailed by physical states. That is, one might have good reasons for thinking that phenomenal properties are conceptually supervenient on physical properties without having the slightest clue for any given phenomenal property what physical or functional properties entail it and why.[18] I have offered a number of such reasons in this paper, to be weighed, of course, alongside other good reasons pro and con. We would then be left with an explanatory gap, to use Joseph Levine's phrase.[19] But it would be a gap that is open in practice not in principle. It would be closable in practice in an individual case by coming to see a conceptual path from physical to phenomenal.[20]

---

[1] I follow Chalmers and recent practice in using 'consciousness' throughout to refer to phenomenal consciousness. All page references are to The Conscious Mind (Oxford University Press, 1996).

---

[2] He describes this view as panprotopsychism, and as the Russelian view. See e.g. 135-6, 143, 154-5.

[3] Chalmers works with the 'no mental difference without a physical difference' type of formulation (33). But I do not think this affects the discussion. Jaegwon Kim persuasively argues that the two formulations are equivalent under certain natural assumptions in his Supervenience and Mind (Cambridge: Cambridge University Press, 1993) (hereafter: SM) 81-2.

[4] Terence Horgan argues that this formulation of supervenience reduces to global supervenience in his "From Supervenience to Superdupervenience: Meeting the Demands of a Material World" Mind 102 (1993) 571-2. It certainly entails global supervenience as global supervenience can be represented as a special case of local supervenience. The entailment from global to local is more contentious.

[5] This position is advocated e.g. by Brian Loar "Phenomenal States" Philosophical Perspectives 4: 81-108, and (second version) in Block, Flanagan and Guzeldere (eds.) The Nature of Consciousness (Cambridge, Mass.: MIT Press, 1997) 597-616.

[6] The denial that consciousness conceptually supervenes on the physical appears as a component of the view known as emergentism that can be traced back to C.D. Broad and J.S. Mill. For a historical survey see e.g. Brian McLaughlin "The rise and fall of the British emergentists" in Beckermann, Flohr and Kim, eds. Emergence or Reduction? (Berlin: De Gruyter, 1992)

[7] Donald Davidson appears to deny the supervenience of any of these three kinds (strong supervenience) of intentional properties on physical properties. He famously argues that there can't be any psychophysical laws, and when pressed by Kim and others who argue that strong psychophysical supervenience entails the existence of psychophysical supervenience laws, he responds by saying that the kind of psychophysical supervenience thesis he endorses is a weak

one and not strong enough to license psychophysical laws of any modality. See his "Thinking Causes" in Mele and Heil (eds.) Mental Causation (Oxford: Clarendon Press, 1993).

[8] See his Naming and Necessity, excerpt reprinted in Rosenthal (ed.) The Nature of Mind (Oxford: Oxford University Press 1991) (hereafter: NM) p 246.

[9] Ned Block and Robert Stalnaker in "Conceptual Analysis and the Explanatory Gap" (forthcoming) appear to deny the conceptual supervenience of a number of nonmental properties on the physical, including biological properties such as being alive. Alex Byrne offers a thoroughgoing denial of the conceptual supervenience of special science properties on the physical in his "Cosmic Hermeneutics" Philosophical Perspectives 13, 1999.

[10] See, e.g. Kim SM essays 14, 15, and 17.

[11] This is reminiscent of Locke's argument, except that he focuses on thought not sensation. Essay Concerning Human Understanding Book IV, Chapter X, Section 10.

[12] Edward Titchener Outline of Psychology (1896) quoted by E.G. Boring Sensation and Perception in the History of Experimental Psychology ,
(New York: Appleton-Century-Crofts, 1942), 10.

[13] See e.g. Hilary Putnam Representation and Reality (Cambridge, Mass.: MIT Press, 1988).

[14] A passage in which Chalmers says that the possibility of inverted spectra establishes a conclusion strictly weaker than the possibility of zombies (101), might lead one to think that he takes the impossibility of spectrum inversion to entail the impossibility of zombies. But he informs me he intended to imply only that the possibility of inverted spectra doesn't entail the possibility of zombies.

[15] For example, the epiphenomenal nature of colour qualia would not be thought troubling on the view that there are widespread logical constraints which nonetheless leave open the logical possibility of spectrum inversion.

[16] These objections to the metaphysical or natural supervenience of sensory properties on the physical apply also to the supervenience of intentional properties, thus contributing support to

---

the general claim that all mental properties conceptually supervene on physical properties. However, if natural or metaphysical supervenience is proposed for all mental properties, this would diminish the force of the objection that they have a very small job to do.

[17] A position of this sort is espoused by Sydney Shoemaker, who takes conceptual constraints to rule out zombies, but holds that consciousness metaphysically supervenes on the physical.

[18] This is akin to Thomas Nagel's claim that we might have some reason to believe mental items are physical items without being in a position to understand how.
"What is it like to be a Bat?" reprinted in NM 427.

[19] "Materialism and Qualia: The Explanatory Gap" Pacific Philosophical Quarterly, 64,4 (1983) 354-61.

[20] I'd like to thank Katalin Balog, Alex Byrne, David Chalmers, Bennett Helm, Brian Loar, Barry Loewer, and Glenn Ross for helpful discussion of earlier drafts of this paper, as well as those who commented on aural presentations of versions of this paper -- none of whom are at all responsible for the faults that remain.

Department of Philosophy
Barnard College, Columbia University
3009 Broadway,
New York, NY 10027-6590
USA