Forthcoming in a volume dedicated to Jaegwon Kim,
edited by Terry Horgan and David Sosa, MIT Press

Phenomenal Externalism, Lolita, and the Planet Xenon

It is usually supposed that the term 'phenomenal character' cannot be defined in non-phenomenal terms. To explain the meaning of the term, one typically begins by saying something non-reductive of the following sort: the phenomenal character of an experience is what it is like subjectively to undergo the experience. Then one proceeds to examples. There is something it is like to feel a tickle in an elbow, experience an itch on one's nose, smell the salty sea air, taste vinegar, have a visual experience of bright red, dangle one's fingers in running water, feel thirsty, experience fear, feel elated. These experiences vary in what it is like to undergo them. Thereby, they differ in their phenomenal character.

Experiences not only have phenomenal character. In many cases, it is noncontroversial that they also carry information, that they tell us things about ourselves or the world around us. Visual experiences purport to inform us as to the colors and shapes of things in our environments; pain experiences signal bodily damage. The informational aspect of experiences is something that many philosophers suppose is entirely separable from their phenomenal character, as indeed is anything external to the experience itself. On this view, what matters to the phenomenal 'feel' of an experience is only how it is intrinsically. Duplicate the causal relations the experience stands in, the cognitive responses the experience generates, the informational links between the experience and other things outside it and you need not thereby have duplicated the experience. It is in principle possible that all these external things are present and yet there is no internal state with phenomenal character at all. This is the so-called absent qualia hypothesis (Block 1980).

Another way to help explain the notion of phenomenal character is to reflect on the famous inverted spectrum hypothesis—the hypothesis, that is, that what it is like for you when you see red things is the same as what it is like for me when I see green things and vice-versa, with corresponding inversions for the other color experiences, even though you and I function in the same ways in color tests and in our everyday behavior towards colored things. Whether or not this hypothesis is true, it can be used to focus our attention on the phenomenal character of an experience just as the description, "the man drinking champagne" can be used to single out a person who in actual fact is female and drinking water (Donnellan 1966).

A further way to fix the referent of the term 'phenomenal character' is to say that it is what gives rise to the explanatory gap (Levine 1983). Tell me everything you like about what goes on physically and functionally in someone who is experiencing red and, it seems, you still won't have told me what it is *like* to experience red. For even after I have all the relevant physical and functional information, I can still intelligibly ask, "Why do those physical and functional goings-on generate *that* phenomenal character (the phenomenal character of the experience of red)? Why *couldn't* another phenomenal character be present?"

I shall call those philosophers who hold that phenomenal character supervenes on internal constitution (where the term 'phenomenal character' is understood as explained above) "phenomenal internalists." On this view, it is metaphysically impossible for intrinsic duplicates to differ with respect to the phenomenal character of their internal

states. Until recently, phenomenal internalism was almost universally accepted in the philosophy of mind. With the rise of representationalism, however, the view has begun to feel some pressure (Byrne 2001; Dretske 1995; Lycan 1996; Jackson 2002; Tye 1995, 2000). For if the phenomenal character of a state is (or supervenes on) its representational content (or *a* representational content the state has that meets certain further conditions, e.g., with respect to the functional role it plays[1]), and content brings in external factors, then prima facie it *is* metaphysically possible for intrinsic duplicates to differ phenomenally.

I say ‘prima facie’ here since there a weaker form of phenomenal externalism which has it that even though phenomenal character is partly constituted by causal relations to external entities, other downstream factors enter into phenomenal character that preclude internal duplicates from differing phenomenally when the normal causes of, or normal environments for, their internal states differ. [2] This form of externalism is compatible with phenomenal internalism, as presented above, and for the purposes of this essay it may be classified as falling within the internalist camp.

In this paper, I want to try to increase further the pressure on the phenomenal internalist by describing a possible case which, it seems to me, is a counterexample to phenomenal internalism, on the assumption that physicalism is true. Having discussed the case, I shall then examine various reasons that have been given for being a phenomenal internalist. I shall argue that all of them fail. In this connection, in Section IV, I shall also try to show that, independent of the opening thought experiment, we have good reasons for holding that the phenomenal character of an experience is not an intrinsic property of it. I shall conclude by discussing some ways in which the thesis of phenomenal internalism might be revised. I shall suggest that none of the revisions result in a viable internalist thesis.

One reaction the phenomenal internalist might have to my argument is to say: so much the worse for physicalism.[3] As will become clear later, this reaction requires the phenomenal internalist to be a dualist not only about phenomenal character but also about the representational properties of mental states. Dualism of any sort is problematic; dualism this broad doubly so. In the present context, however, my aim is not to refute the dualist. I am content to show here that a choice needs to be made: either give up phenomenal internalism or give up physicalism.

Given physicalism, the thesis of phenomenal internalism reduces to the thesis that it is metaphysically impossible for microphysical duplicates to differ with respect to the phenomenal character of their internal states. To refute this thesis, it suffices to produce an example of two entities that are microphysical duplicates in some metaphysically possible world *W* without the two entities being phenomenal duplicates in *W*. It is to the task of constructing such an example that I turn in sections I and II.

---

[1] I put it this way, since pluralism is a viable option with respect to the representational content of experience.

[2] See, e.g., Colin McGinn on what he calls “weak externalism” in his 1989.

[3] This might be the reaction of Jaegwon Kim. Although Kim does not take up the issue of internalism versus externalism directly in his 2005, it is clear from his remarks (pp. 172–3) that he remains a phenomenal internalist. It is also clear that he rejects physicalism for individual qualia or phenomenal characters. However, my argument creates trouble for Kim, since he is not a dualist about the representational properties of mental states.

I. The tree with many ‘brains’

On the planet Xenon, there are massive trees. These trees produce many very large, hanging pods every four years. The pods grow gradually and depend for their development upon the copious rainfall that is found on Xenon. When the contents of the pods are ready for harvesting, their shells begin to crack open. This process is hastened by the many electrical storms that occur. Lightning often strikes the trees and the electricity is conducted throughout the tree limbs and into the bodies of the pods, thereby causing them to split apart once they have grown to a certain size.

The contents of the pods are eaten by the people who live on Xenon (Xenonites, as I shall call them). Sometimes, the Xenonites are so anxious to eat the pod contents that they go out during the storms and devour them straight from the trees as the pods split open. Xenonites are very different from us. Their physiology is not brain and spine based as ours is. They do not have any neurons inside their bodies. Instead, their minds function hydraulically.

The pods themselves are each large enough to contain a human brain and remarkably their contents, just before harvesting, are chemically very like human brains in which there is no activity. Even more remarkably, one particular pod (call it ‘XP1’), during an electrical storm that infuses it with electricity for fifteen minutes, is actually a microphysical *duplicate* of an active brain—one belonging to a human being named ‘Lolita’ who for the same period of time on Earth initially is having sexual intercourse, and then (after eight minutes) smoking a cigarette and drinking green chartreuse.

I take it to be obvious that the pod contents are not themselves genuine brains. Upon rare occasion, they may briefly be microphysically identical to live human brains, as with XP1 for fifteen minutes, but brains they are not. In making this assertion I am not supposing that there cannot be brains without bodies. Clearly, there can be brains in vats or brains removed from bodies which are subsequently destroyed. However, in these cases, it is at least true that the brains were *designed* to control the bodies of the creatures to which they belong or did belong, even if those creatures no longer exist or have had their usual bodies taken from them. Nor do I wish to deny that swamp brains are possible; intuitively, a molecule by molecule duplicate of my brain that is functioning for a sufficient period of time *as* a brain inside a human body is a brain even if it was accidentally created by a chemical reaction that took place in a swamp.

What I am denying is that something becomes a brain *simply* by replicating a brain microphysically. We don’t suppose that something becomes a key just by replicating a key microphysically. Think, for example, of a plastic card that happens to replicate a card key, but is actually designed for use as a credit card for Xeroxing articles at a library. Nor do we suppose that something becomes a tiger just by being a microphysical duplicate of a tiger. If the thing has the wrong evolutionary history, it isn’t a tiger no matter how closely it resembles one. A voltmeter, as all will agree, doesn’t become a speedometer unless it is designed (or at least used) to measure speed. An intrinsically identical voltmeter used to measure volume may be a fuel gauge instead. Similarly, a microphysical duplicate of my little finger that is actually a toe on a three-headed Martian is not also a finger. The same points apply *mutatis mutandis* to microphysical duplicates of human brains.

XP1, then, is not a brain. It was not designed by nature to function as a brain; nor has it become a brain by taking on the appropriate control role with respect to a body. By

pure accident, XP1 briefly replicates a brain microphysically, but that is not enough to make it a brain.

Does XP1, for the period of time during the storm in which it is microphysically identical to a particular human brain, undergo experiences, all of which are phenomenally identical to the experiences of the relevant human on Earth? I say *no*.

II. Lolita, XP1, and bodily sensations

Consider the bodily sensations Lolita undergoes. There is a locational component to these sensations. When she feels a pleasurable tickling sensation in her upper arm, say, the bodily location she experiences for the tickle contributes to, or is at least fixed by, the overall phenomenal character of her sensation. This is not to imply that bodily experiences that have the same phenomenal character must represent the same bodily part. There is no obvious reason to deny that a creature might feel a pain in a finger that is shaped just as one of my toes is and which feels to the creature, location-wise, just as a pain in a toe does to me. This seems possible since the fingers of such a creature might bear the same torso-relative bodily locations as my toes. Even so, in such a case, there is a commonality in represented torso-relative bodily location notwithstanding the difference in objective, bodily part location; it is the former that, I claim, contributes to, or is at least fixed by, phenomenal character. The relevant location is *represented* location, since one can feel a pain in a given bodily location even though one has no bodily part in that location (as, for example, if the appropriate bodily part has been amputated).

Here is another way to make the point. Suppose I feel a pain in a finger and I move my finger to a different location relative to my torso. Then my pain feels to be in a different location and this entails that there is a difference in phenomenal character before and after the movement. By contraposition, then, sameness in phenomenal character entails sameness in felt torso-relative location. Since sameness in felt location necessitates sameness in represented location, sameness in phenomenal character in this case necessitates sameness in represented torso-relative location.

In making these remarks, I am not assuming the truth of representationalism either with respect to phenomenal character generally or more narrowly with respect to the phenomenal character that attaches to the experience of bodily location. According to representationalism in its weakest form, necessarily experiences that have the same representational content have the same phenomenal character. This is not assumed above even for the special case of bodily location phenomenal character; nor is it a consequence of what I say.[4] As just noted, what my comments entail is only that bodily sensations that feel alike with respect to bodily location (and thus have the same locational phenomenal character) must represent the same torso-relative bodily location.[5]

---

[4] For a detailed discussion of bodily sensations (and pain in particular) from a representationalist perspective, see my forthcoming 2006.

[5] I shall not press the point here since it is not needed for present purposes, but, in my view, our experiences generally have what might be called a "presentational phenomenology." For the appropriate external aspects, experiences with the same phenomenal character present the same aspects of the world to us or the same aspects of our bodies (or sometimes both). I focus on the case of phenomenal location, since it seems especially clear-cut and compelling.

Consider now XP1. If XP1's bodily sensations, if any it has, feel to it just as Lolita's bodily sensations feel to her, then XP1 must have an internal state representing a bodily part with the same torso-relative location as Lolita's arm, given that Lolita has a tickle in her arm. But this seems *very* implausible. XP1 does not belong to a species of creatures with arms (or bodily parts that are positioned relative to torsos in the ways that arms are). It is not a brain belonging to such a species. Indeed, as we saw earlier, XP1 is not a brain at all. Further, XP1 does not have any internal states that are *supposed* to indicate arms or disturbances in arms (or arm-like parts).[6] Nor does XP1 have any internal states that causally co-vary with disturbances in, or on the surface of, arms; for XP1 is hanging in a pod on a tree. Also, XP1 has no internal states that lead to arm-rubbing or arm-moving behavior. To be sure, XP1 has internal states that *would* causally co-vary with arm disturbances (or would lead to arm-related behavior), *were* XP1 connected in a certain complicated way to an appropriate human body for the crucial fifteen minutes. But why suppose that this is relevant to what, if anything, the internal states of XP1 represent, as it hangs on the tree in a pod? After all, there are many other possible bodies XP1 could be connected to in complicated ways so that the same internal states of XP1 would then causally co-vary with disturbances in regions of XP1's torso.

What, then, could make it the case that XP1 has any internal states representing arms (or bodily parts spatially related to torsos as arms are)? The physicalist has no satisfying answer to this question. Admittedly, she might insist that there is an answer, but not one that we can grasp or, at any rate, yet formulate. But this seems very unsatisfying. Prima facie, in the case of XP1, there simply isn't any physical fact that can ground the relevant representational relations.

Of course, the dualist might respond that there are *primitive* representational relations that XP1's internal states bear to arms or arm-like bodily parts. These relations are non-physical and it is just a brute fact that they obtain for any intrinsic duplicate of a being with internal states representing arms or arm-like parts. This view seems to me extremely implausible but, as noted in the opening part of the essay, my aim here is not to refute the dualist.

Given physicalism, then, the conclusion to which we are led is that XP1's internal states do *not* represent arms (or bodily parts spatially related to torsos as arms are). XP1's bodily sensations, if any it has, do not feel to it just as Lolita's bodily sensations feel to her. It follows that what it like for XP1 is not overall the same as what it is like for Lolita.

Compare the Xenon tree case with the case in which a mad neurophysiologist of the future probes the brain of Lolita while doing neurosurgery and causes her to go into just the brain states she would have gone into had she been engaging in sex, and then drinking green chartreuse and smoking a cigarette. In this case, I have no hesitation in saying that Lolita is experiencing just what she would have experienced, had she really been doing these things. What is the difference between this case and the Xenon one? At the most general level, the difference is one of background context. In this case, there is a normal setting, relative to which the given situation counts as abnormal, and this setting can be used to justify the attribution of experiences of just the phenomenal type that would have been present in the normal case. In the Xenon tree case, there is no normal setting that can play this role.

---

[6] For ease of exposition, for the rest of this paragraph, I leave out the parenthetical qualification.

But what if some visiting earth scientists had placed XP1 in an appropriate human body for the crucial fifteen minutes? With the right connections, XP1's behavior, both verbal and nonverbal, would have mirrored that of Lolita. Isn't that evidence that XP1 is psychologically like Lolita?

My reply is that XP1 could have been hooked up to many other possible non-human bodies and thereby have produced very different narrow verbal and nonverbal behavior. This being so, there is no clear reason to hold that the scenario in which XP1 is suitably connected to a brainless human body shows anything about the psychological life of XP1, as it hangs from the tree in a pod. Consider a card designed for use as a credit card for Xeroxing (to return to the earlier example). It could have been used for all sorts of things. For example, it could have been used by a thief to open hotel room doors. It hardly follows from this mere possibility that the card *is*, in actual fact, a key card.

Suppose now that the original scenario had been different and that just as the crucial fifteen minute period was ending, XP1 had been placed in a live human body and connected to it just as our brains are standardly connected to our bodies. Would not the embodied individual remember immediately afterwards having had just the very sensations Lolita had? If so, then isn't the best explanation of these memories that XP1 in this case really did have the relevant experiences? And if this is correct, then shouldn't we agree that XP1 in the original case had those experiences too, especially since the difference between the two cases occurs only *after* the fifteen minutes during which XP1 is infused with electricity? How can a later difference make a difference to what occurs before?

Let us grant for the moment that the being with a human body has apparent memories of some earlier experiences. Then we should also grant that this being has other apparent memories. She seems to remember lying on a bed, smoking a cigarette, talking to another person, drinking chartreuse, having her arm tickled. In reality, none of these things took place. Her beliefs about her past real-world life are false, as are her beliefs as to who she is. Furthermore, her beliefs about experiences she had prior to the crucial fifteen minute period are false too. Given all this, it is not in the least obvious that the best explanation of her beliefs about her most recent experiences are in any better shape. There seems no obvious rationale, other than a blind adherence to phenomenal internalism, for treating these beliefs as any different from her other beliefs about her past. After all, there is no privileged access to past psychological states any more than there is privileged access to past, objective goings on. Surely, the simplest, most coherent view is that her beliefs about the past stand or fall together.

It is also worth noting that it is not *obvious* that a real psychological subject exists immediately after XP1 and the human body are joined together. Of course, it appears to others that there is such a subject, but appearances can be deceptive. Something can appear to be a key without being one; something can appear to be a tiger and be something else; something can appear to be a finger when it is actually a toe; something can appear to be a brain and not be a brain. Why cannot the same be true of psychological subjects? Given the complexity of the resulting structure, and the physical similarity between its parts and those of real psychological subjects, it is tempting to assume that the structure *must* be a psychological subject. But this would be hasty. Some further argument is needed.

I shall not pursue this point here. I merely observe that *if* the combined structure is not a psychological subject, or at least is not such a subject *initially*, then XP1 prior to

embodiment is not a subject either. [7] But if XP1 is not a subject, then XP1 cannot have *any* experiences prior to embodiment. Experiences cannot exist *un*owned any more than laughs can exist *un*laughed or screams can exist *un*screamed. For each experience, there must be an experiencer—someone for whom there is something it is like. But if XP1 has no experiences, then there is *nothing* it is like for XP1 at all.

Returning now to my earlier claim that there is a phenomenal difference between XP1's experiences and those of Lolita, a further worry may surface. For all that I know a priori, might I not *myself* be in a pod now on Xenon, deluded into believing, on the basis of inaccurate experiences, that I am in Texas, sitting at a desk with a computer before me, my legs crossed and my arms resting on the desktop as I type? It seems so, in which case it is conceivable that there is a pod content on another planet that has experiences phenomenally exactly like some particular human being (namely, me). But if this is possible, then does it not bring into question my claim that XP1 on Xenon has experiences phenomenally *different* from those of Lolita?

I think not. The conceptual possibility of my being in a pod on Xenon does not show that this is metaphysically possible. It seems plausible to suppose that I am essentially human and thus that if I were now in a pod on Xenon, it could only be because I had been taken there from Earth without my knowledge, in which case the scenario, though now metaphysically possible, has no bearing on the one in which XP1 grows naturally in a pod in its home environment and lacks any connection with human beings or other species of biological organisms.

In my view, then, the Xenon example provides us with a possible case in which a standardly embodied creature with a brain and a microphysical duplicate of that brain differ phenomenally. It does not yet show that microphysical duplicates can differ phenomenally. However, there is a simple extension of the thought experiment that does. Let the subject on Earth be someone whose body has been destroyed and whose brain has been envatted and subsequently is supplied by inputs from a supercomputer, so that XP1 is now a microphysical duplicate of a person who is systematically hallucinating the act of sexual intercourse followed by smoking a cigarette and drinking green chartreuse. Here we really do have microphysical duplicates that differ phenomenally. The case thus supports phenomenal externalism.

This view will not be shared by everyone. Some will no doubt dig in their heels and insist that XP1 is a brain and that it does have phenomenally identical experiences for the fifteen minute period, however strange that initially seems. Still at a minimum, even those philosophers who react in this way should agree that the example raises a serious doubt as to whether any two microphysical duplicates must be alike phenomenally. Thus, the Xenon tree example *at least* shows that phenomenal internalism is not a position that is self-evidently true or one that cannot reasonably be disputed.

III. Phenomenal internalism

---

[7] Here is more food for thought on the question of whether XP1 is a psychological subject. Suppose that the Xenonites in their zeal to taste the pod contents had pulled XP1's pod from the tree before the final electrical storm ended and had eaten XP1. Would they have been doing a *bad* thing? It is tempting to suppose, that, even if consequentialism is true, no question of goodness and badness arises here. However, if XP1 is a psychological subject, the Xenonites' behavior is open to moral assessment.

So, why be a phenomenal internalist? To this question, I maintain, there is no good answer.[8]

The first reason that might be offered is that the immediate causes of phenomenal states are neural states. There is a direct causal dependence between the phenomenal and the neural. So, of course, if you fix what goes on in the brain, you fix what goes on at the level of phenomenology.

The obvious reply to this is that causal dependence does not establish metaphysical dependence. If it did, externalism about belief contents would be refuted by the same reasoning too. And nobody believes that.

A second reason is that empirical work on color vision traces the phenomenology of color experiences to opponent processing channels in the brain. Differences in color experience phenomenology are explained by reference to different activation levels in these neuronal channels (Pautz forthcoming 2006). So, again if you fix the brain events, you fix the phenomenal character, or at least the phenomenal character of color experience.

My reply is that explanatory dependence does not establish metaphysical dependence. Consider the hypothesis that the fire started *because* of the short circuit. Suppose that this hypothesis is true. The fact that an identical short circuit in a different setting did not start a fire is no objection to the proposed hypothesis. For explanation is relative to a background context. The short circuit explains why there was a fire, but the short circuit alone does not metaphysically necessitate a fire. Other factors extrinsic to the short circuit are relevant. Analogously, the phenomenal externalist claims that, for phenomenal character, the relevant factors include matters outside the brain.[9]

A third reason is provided by the observation that I myself might conceivably be a brain in a vat, wired to a computer that is supplying me with inputs. This being so, how things are outside the brain does not really matter to phenomenology. If the brain states are the same, the phenomenology must be the same, whatever the external causes (Horgan and Tiensen 2002).

This is a non sequitur. The externalist can grant that it is metaphysically possible that human brains in vats have experiences. The externalist can also grant that it is conceptually possible not only that I am a brain in a vat but also that I am not human at all. However, given that, in fact, I am human, it is metaphysically necessary that I am human. So, it is metaphysically necessary that if I am a brain in a vat, I am a human brain in a vat. So, any metaphysically possible scenario involving me is one in which certain external facts obtain. To assume that I could have just the phenomenal life that I actually have, *whatever* my external setting, is either to assume a conceptual possibility that cuts no ice against the externalist thesis that it is metaphysically possible for some microphysical duplicates to differ phenomenally or it is to assume that something is metaphysically possible which in reality is not. Either way, there is no trouble for phenomenal externalism.

A fourth reason (suggested to me by Cory Juhl) appeals to causal considerations. Consider a microphysical duplicate of the present time slice of our world (call it '*MD*') that a) is governed by the same physical laws as our world and b) is the initial slice of a world, *W*, with no history prior to the present time. If the physical world is causally closed, then *W* will unfold physically just as the actual world will. So, future behavior in

---

[8] I am not alone in taking this view. See e.g. Lycan, op.cit.

[9] For more on explanation and metaphysical necessitation, see Byrne and Tye, 2006.

W will be the same as in our world. Now, given that phenomenal externalism imposes some sort of backward-looking tracking requirement on phenomenal character, since *MD* is the first time slice of *W*, there is no phenomenal character tokened in *W* at the present time. *MD*, then, is a zombie replica of the current time slice of the actual world. But if this is the case, then the presence of phenomenal character in our world makes no difference to future behavior. In short, phenomenal externalism make phenomenal character epiphenomenal, and that is unacceptable.

This is unconvincing. It is certainly true that *W* will unfold *microphysically* just as the actual world will; but it does not follow from this that *W* will unfold at *higher* levels in the same way. Note first that *W* lacks any human beings, since there are no beings in *W* with an evolutionary history that matches that of actual humans. To be sure, the beings who inhabit *W* at the present time are microphysical duplicates of actual humans, but even so they are not human. Not only will there be gross biological differences but there will be other teleo-functional differences that engender differences in propositional attitude contents and intentional behavior in the two worlds (on any partly backward-looking externalist accounts of these). So, even if subsequent bodily movements are the same in *W* as in the actual world, there is still plenty of room for phenomenal character to be causally efficacious both with respect to propositional attitudes (for example, beliefs) and with respect to intentional behavior.

The key point here is that the above example presents no *special* problem for the externalist. Essentially the same objection could be raised just as easily with respect to the causal efficacy of belief contents. And the reply by the content externalist would take the same form as the one I have just given.

A final reason that might be offered for being an internalist about phenomenal character is that the phenomenal character of an experience is an intrinsic property of it. Given that this is so, of course intrinsic duplicates must be alike phenomenally.

Too fast, I respond. How should we understand the term 'intrinsic' in the premise here? The term 'intrinsic' sometimes means *essential*. Take the visual experience I am undergoing now, as I view the page before me. It is not implausible to hold that this experience could not have had a different phenomenal character. If I had been having a visual experience with a different phenomenal character, then it would not have been *this* very experience.[10] If the phenomenal character of my experience is essential to it, then its phenomenal character is intrinsic to it in the above sense. If this is how we understand what it is for a property to be intrinsic, then the argument of the qualia internalist is straightforwardly invalid.

If this is not obvious, consider the property of being caused by a mosquito. This, let us grant, is an intrinsic property in the above sense of a mosquito bite. Patently it does not follow from this that individuals who are intrinsic duplicates must be alike with respect to whether they have skin punctures that have been caused by a mosquito

There is another sense of 'intrinsic', however, that contrasts with extrinsic. The recent literature on the metaphysics of intrinsic properties shows that it is not easy to say in full detail just what an intrinsic property is in this second sense.[11] But, at least for our purposes, the following remarks will suffice:

---

[10] I myself am disinclined to accept this claim. See Tye 2003, chapter 4.

[11] A summary of this literature is provided by Brian Weatherson in his entry "Intrinsic versus Extrinsic Properties" for the Stanford Encyclopedia of Philosophy.

> An intrinsic property is "a property a thing has (or lacks) regardless of what is going on outside of itself." (Yablo 1990)
>
> "The intrinsic properties of something depend only on that thing; whereas the extrinsic properties may depend, wholly or partly, on something else." (Lewis 1983)
>
> "If some thing has an intrinsic property, then so does any perfect duplicate of that thing; whereas duplicates situated in different surroundings will differ in their extrinsic properties." (Lewis 1983)

Thus, being a sister is extrinsic, as is being 3 feet from a table. By contrast, having mass is intrinsic as is being round (unless the shape of a thing can be made to vary with the curvature of the space in which it is located).

Given the above understanding of an intrinsic property, once again it does *not* follow simply from the assumption that the phenomenal character of an experience is one of its intrinsic properties that individuals who are intrinsic duplicates cannot differ phenomenally. To see this, consider the state of shaking hands with a stranger. Touching something else is an intrinsic property of this state; but individuals who are intrinsic duplicates could still be such that one is shaking hands with a stranger (and thus is touching something else) and the other is not.

Furthermore, and more importantly, why should we now accept the assumption? One reply is that the truth of the assumption is revealed by introspection. But this is not so. While it is true that introspection does not reveal phenomenal character to be an extrinsic property, this is perfectly compatible with supposing that it is. Of course, those who accept the thesis of *revelation* with respect to phenomenal character—that the nature of phenomenal character is fully exposed in introspection—will balk at this reply. But the thesis of revelation is a philosophical thesis, not a piece of commonsense. It is no more to be trusted in this context than it is to be trusted with respect to the nature of color. The fact that when one sees something red and it looks red to one, redness does not look to have a hidden nature shows nothing with respect to the color red. Red itself does not look any way when one sees something red.[12] Correspondingly, the phenomenal character of my current visual experience, for example, is not presented to me, via introspection, as being an intrinsic property or as being an extrinsic property. The nature of phenomenal character is left open.

Admittedly, I have not shown in giving this response that the phenomenal character of an experience is *not* an intrinsic property of it, in the second sense of the term 'intrinsic'. So, the phenomenal internalist might respond that she is entitled to assume that it is until good reasons are offered for supposing otherwise. However, it seems to me that this response places the phenomenal internalist on very shaky ground (even leaving to one side the thought experiment with which I began); for what now *motivates* this assumption? After all, it is not an assumption that is required by the thesis of phenomenal internalism. That is to say, phenomenal internalism is compatible with the view that the phenomenal character of an experience is an extrinsic property of it.

In the next section, I pursue the issue of intrinsicality by presenting an argument that phenomenal character is not, by its nature, intrinsic (in the second sense of 'intrinsic' distinguished above). The conclusion of this argument, if true, does not establish the truth of phenomenal externalism, but it leaves the phenomenal internalist without any

---

[12] What looks some way is the surface which is red.

basis whatsoever (so far as I am aware) for being a phenomenal internalist. Further, *one* possible explanation, of course, of why the phenomenal character of an experience is extrinsic to it is that it is an environment-involving property, as the phenomenal externalist supposes.

IV. More on whether phenomenal character is intrinsic

In what follows, I shall call any intrinsic property of an object that is either a microphysical property of that object or a property that is metaphysically necessitated by an intrinsic, microphysical property of that object a "P-property". Consider, then, a very simple *token* visual experience *v*—the experience of a flash of light at time *t*, say. I begin with something that is undeniably true:

(1) If the phenomenal character of *v* is an intrinsic property of *v* then either it is a P-property of *v* or it is an intrinsic, irreducibly nonphysical property of *v*.[13]

From the definition of an intrinsic property, we have

(2) Microphysical duplicates situated in different surroundings do not differ in their P-properties.

Next, I assume token physicalism with respect to *v*:

(3) *v* is a neural event (or state token).

Obviously, this premise is not one that will be universally accepted. But, as I noted at the outset, I assume here physicalism. One reason for doing so is that if v is a nonphysical event, then there cannot be a complete neurophysiological explanation of bodily behavior; and the current evidence from neurophysiology makes it very likely that, in principle, there is. This is sometimes called "the shadow cast by neurophysiology."

My final premise is as follows:

(4) A microphysical duplicate of *v* in a petri dish has no phenomenal character.

The reasoning behind (4) is simply this. Suppose that there is a microphysical duplicate of *v* in a petri dish. This duplicate will be a certain connected structure of firing patterns in an appropriate group of neurons in the dish. However, there won't be any token experience in the petri dish. For patently there is no subject in the dish to have an experience and, as noted in section I, experiences cannot exist unowned. But if there is no experience in the dish then there is no entity in the dish having phenomenal character. And if this is true, then (4) follows.

One objection to this reasoning is that the relevant structure of neuron firings will be widely scattered, bringing in events from many different regions of the brain, and thus will not obviously be the sort of event that could occur in a petri dish (even an oversized one) or without a subject for that matter. This seems a very implausible view, however, for such a simple visual experience as that of a sudden flash of light. Further, it confuses

---

[13] Some philosophers are prepared to countenance nonphysical properties that are nonetheless metaphysically necessitated by the microphysical. I deny that there are any such qualities, since I deny that there are brute supervenience laws of the sort needed by this view. (For more here, see Tye 1995, chapter 2). But leaving this to one side, the argument of the present section assumes an understanding of 'irreducibly nonphysical' under which such properties do not count as irreducibly nonphysical.

the question of which token physical event is the token experience *v* with the question of which overall physical setting must be present for that physical event to have the phenomenal character of *v* or to be an experience at all.

Here is a parallel. Suppose I launch the rocket by pressing the red button at time *t* in Mission Control. My launching the rocket at *t* just is my pressing the button at *t*.[14] But what makes my pressing the button a rocket launching is something involving many other events. This is why a microphysical duplicate of my button pressing located against a different background need not be a rocket launching.

Correspondingly, it is certainly the case that without many brain events occurring at *t*, including activity in the brain stem, there would be no visual experience with the phenomenal character of *v*. But MEG scans reveal sudden localized activity in the mesial occipital cortex temporally coinciding with v. *This* token physical activity, the physicalist should say, is *v*. It has the right temporal length and it plays the right causal role. The other events form part of the background without which this activity would not have the psychological character of *v*.

One possible reply to this is to say that *v* is not an event at all, but rather a state token and that the best candidate brain state token for identification with *v* (or for constituting *v*) is a global one—that of the given brain's having such-and-such activity in the brain stem and so-and-so activity elsewhere, including activity in the mesial occipital cortex. This seems ad hoc, however. Why suppose that there is such a complex state token? To be sure, there is a relevant, very complex neural property the subject instantiates at *t*, that of having a brain with activity in regions *X*, *Y*, *Z*, etc. And there is also a corresponding complex neural property the brain instantiates. But what reason is there to hold that there is a state *token* that lasts just as long as *v* and has its causal powers? None that I can see.

From (1)–(4), we may conclude

(5) The phenomenal character of v is not a P-property of *v*.

So,

(6) If the phenomenal character of v is an intrinsic property of *v*, then it is irreducibly nonphysical.

But if the phenomenal character of *v* is irreducibly nonphysical, then it is hard to see how it can be causally efficacious with respect to behavior. Since patently it is, we have

---

[14] This assumes a sparse view of token events of the sort persuasively argued for by Davidson (1970). The claim that my pressing the button at *t* just is my launching the rocket at *t* is not undermined by the thought that I could have pressed the button without the rocket being launched whereas I couldn't have launched the rocket without the rocket being launched. What this shows is that the property of being a rocket launching is only an accidental property of that particular button pressing (the very event that is, in fact, a rocket launching). This is on a par with the claim that inventing bifocals is only a contingent property of Benjamin Franklin (the man who, in fact, invented bifocals).

Another response here in the general spirit of Davidson is to say that my pressing the button at *t* is my launching the rocket at *t*, even though the relationship is not one of strict identity. The former event *constitutes* the latter; and constitution does not require possession of all the same modal properties. This is my preferred view of the relationship between token visual experience v and the relevant neural event.

(7) The phenomenal character of *v* is not irreducibly nonphysical.

This also follows trivially from the thesis of physicalism. By (6) and (7), we reach the conclusion:

(8) The phenomenal character of *v* is not an intrinsic property of *v*.

So, it is not the case that the phenomenal character of an experience is, by its very nature, an intrinsic property of the experience.[15]

V. Possible revisions to the internalist thesis

Instead of holding that intrinsic duplicates cannot differ phenomenally and thus that microphysical duplicates cannot differ phenomenally, if physicalism is true, the internalist might now propose the following:

(BT) *Brains* that are microphysical duplicates cannot differ phenomenally.

One immediate question that arises in connection with (BT) is whether the property of being a brain is an intrinsic property (in the sense of 'intrinsic' that contrasts with 'extrinsic'). For reasons given in Section I, the answer seems clear: it is not. What, then, makes (BT) true (if it is true at all)?

The answer must be that it is the possession of the extrinsic property of being a brain or some other extrinsic property common to all possible brains that makes microphysical duplicates having that property phenomenal duplicates. This is because if the property, the possession of which determines that microphysically identical brains cannot differ phenomenally, is intrinsic, then it will be possessed by all those entities that are not brains but that are microphysical duplicates of brains. In that case, it will be necessarily true that any two entities that are microphysical duplicates and that possess the relevant intrinsic property common to all possible brains are phenomenal duplicates. The reasoning of Sections I and II, in connection wit the Xenon tree thought experiment, shows that there is no such phenomenally determinative intrinsic property (or at least brings the view that there is into serious doubt).

However, the property of being a brain obviously is not a plausible candidate for the role of extrinsic property that guarantees phenomenal sameness in microphysical duplicates that are brains. So what is the crucial extrinsic property? It is, I suggest, radically unclear. Moreover, to suppose that there is such a property is to place an emphasis on the extrinsic that is not really in the spirit of internalism anyway.

Here is another possible internalist thesis:

(BT*) Microphysical duplicates of structures consisting of brains appropriately connected to sense organs cannot differ phenomenally.

In this case, a slightly revised version of the Xenon tree experiment is relevant to showing that the thesis is in trouble.

Suppose now that the contents of the pods on Xenon grow in all sorts of irregular ways and that, as it happens, in one of the pods there is a complex structure (call it 'XP2') that, during an electrical storm is infused by electricity so that, for a brief period of time,

---

[15] The phenomenon of transparency also counts against the view that phenomenal character is intrinsic. For more on transparency, see Moore 1903; Harman 1990; Tye 2002.

it microphysically duplicates Lolita's brain *and* sense organs, along with their connections. Again, I take it that there is no brain inside the pod. Nor, in my view, are there ears, eyes, a nose, etc. There are simply things that accidentally replicate these human sense organs. Of course, were we to be visitors to Xenon from Earth, we might think at least initially that a part of XP2 which was lying on the ground after its shell had split apart really was an ear, say. But if we did think that, we would be wrong. The context here is different from that which obtains on Earth when we confront something that physically duplicates an ear; and this difference makes a difference. Let me elaborate further.

Suppose that, as you are walking home, you notice on the sidewalk something that looks just like the speedometer on your Honda motorbike. Further investigation reveals that it even has the letters 'honda' inscribed on the front and that in all obvious respects it closely resembles a Honda speedometer. You infer that in all probability it *is* a Honda speedometer. This inference is legitimate since, our home environment being the way it is, the thing on the sidewalk is almost certainly *not* a swamp speedometer (something accidentally created by molecules randomly coming together after a huge explosion, say). Rather, it is almost certainly a device designed for a certain purpose.[16]

Now Honda makes not only speedometers but also fuel gauges. Suppose the following scenario obtains: intrinsically, some of the Honda motorbike speedometers are exactly like some of their car fuel gauges. In this case, there is a real question as to whether the thing on the sidewalk is a Honda speedometer, notwithstanding the similarities; for there is a real question as to what the thing on the ground is supposed to do, what it was designed to do. It might be a speedometer, but then again it might be a fuel gauge. There is no way of telling which without further information about the origin of the thing and the intentions of its designers.

In the case of XP2, there is certainly a part that physically resembles one of Lolita's ears. But this part was not designed to respond to sounds. It is not *supposed* to register sounds. Ex hypothesi, it has no such job given to it by Mother Nature. Further, this part of XP2 not only closely resembles an ear but also—let us suppose—closely resembles a benign tumor that sometimes grows in the stomachs of creatures living on another far away planet. Which of these is it? Picking either would be as arbitrary as insisting that a swamp replica of a speedometer is really a speedometer and not a fuel gauge. So, we should pick neither. Thus, the relevant part of XP2 is not an ear. What goes for this part goes for the others. XP2, therefore, is not a structure consisting of a brain to which sense organs are attached.

With this point established, the argument against the truth of (BT*) now proceeds in a parallel way to the earlier argument against the original internalist thesis. If in place of (BT*), the internalist suggests

(BT+) Structures that consist of brains appropriately connected to sense organs and that are also microphysical duplicates cannot differ phenomenally,

the points made in connection with (BT), the first alternative proposal, apply mutatis mutandis. As far as I can see, then, once the original internalist thesis is refuted or at least brought into real doubt, the revised theses are no improvement.

I conclude that phenomenal internalism is a view we have no reason whatsoever to accept and serious reason to question (at least if physicalism is true). It is no more or less

---

[16] Cf. Dretske 1995.

than a dogma, a hangover from an outdated Cartesian conception of the mind. The time has come to take phenomenal externalism *very* seriously indeed.[17]

Michael Tye

The University of Texas at Austin

REFERENCES

Block, N. 1980 "Troubles with Functionalism," in his *Readings in the Philosophy of Psychology*, Volume 1, Cambridge, Mass : the MIT Press.

Byrne, A. 2001 "Intentionalism Defended," *Philosophical Revie*w 110, 190–240.

Byrne, A. and Tye, M. forthcoming 2006 "Qualia Ain't in the Head," *Nous*.

Davidson, D. 1970 "The Individuation of Events," in *Essays in Honor of Carl G. Hempel*, ed. by N. Rescher, Dordrecht, Holland: Reidel.

Donnellan, K. 1966 "Reference and Definite Descriptions," *Philosophical Review* 75, 281–304.

Dretske, F. 1995 *Naturalizing the Mind*, Cambridge, Mass: Bradford Books, the MIT Press.

Harman, G. 1990 "The Intrinsic Quality of Experience," in Philosophical Perspectives, vol. 4, ed. by J. Tomberlin , Atascadero, Calif: Ridgeview.

Horgan, T. and Tiensen, J. 2002 "The Phenomenology of Intentionality and the Intentionality of Phenomenology," in *Philosophy of Mind: Classical and Contemporary Readings*, ed. by D. Chalmers, Oxford University Press.

Jackson, F. 2002 "Mind and Illusion," in *Minds and Persons*, ed. by Anthony O'Hear, Cambridge University Press.

Kim, J. 2005 *Physicalism or Something Near Enough*, Princeton University Press.

Lewis, David 1983, "Extrinsic Properties," *Philosophical Studies* 44: 197–200.

Levine, J. 1983 "Materialism and Qualia: The Explanatory Gap," *Pacific Philosophical Quarterly*, 64, 354–361.

Lycan, W. 1996 *Consciousness and Experience*, Cambridge, Mass: Bradford Books, the MIT Press.

McGinn, C. 1989 *Mental Content*, Blackwells.

---

[17] I would like to thank Chris Peacocke for some helpful remarks.

Moore, G. 1903 “The Refutation of Idealism,” in his *Philosophical Papers*, London: Routledge and Kegan Paul.

Pautz, A. forthcoming 2006 “Sensory Awareness is not a Wide Physical Relation,” *Nous*.

Tye, M. 1995 *Ten Problems of Consciousness*, Cambridge, Mass: Bradford Books, the MIT Press.

Tye, M. 2000 *Consciousness, Color, and Content*, Cambridge, Mass: Bradford Books, the MIT Press.

Tye, M. 2002 “Representationalism and the Transparency of Experience,” *Nous* 36, 137–151.

Tye, M. 2003 *Consciousness and Persons: Unity and Identity* Cambridge, Mass: the MIT Press.

Tye, M. 2006 “Another Look at Representatiomaliam about Pain,” in a collection of essays on pain, ed. by M. Aydede, Cambridge, Mass: Bradford Books, the MIT Press.

Weatherson, B. “Intrinsic versus Extrinsic Properties,” in the online *Stanford Encyclopedia of Philosophy*.

Yablo, Stephen 1999, “Intrinsicness”, *Philosophical Topics* 26: 479–505.