

Roberto Casati

Version of May 07, 2002

To appear in : A. Chatterjee, ed., Towards a Science of Consciousness.

Qualia Domesticated

1. Consider the following argument

- (1) If panpsychism is true, then the hard problem of consciousness is solved
- (2) Physicalism is true
- (3) Physicalism entails panpsychism.

We conclude that

- (4) The hard problem of consciousness is solved.

This is a valid argument, and one whose conclusion has a certain appeal. What about the premisses? How exactly is panpsychism a solution to the problem of phenomenal consciousness? Who can take panpsychism seriously, and how can panpsychism be entailed by physicalism? A little forcing is assumed in suggesting to consider a philosophical argument whose conclusion is panpsychism. But I think the exercise is worthwhile, provided we spell out all the consequences of forcing.

2. The hard problem of consciousness is the problem of understanding the place of qualia, or phenomenal reality, in a world that seems to be completely describable in physical (*hence*, it is alleged, non phenomenal) terms. Here is one formulation of it (Chalmers 1996, p. 5):

“When someone strikes middle C on the piano, a complex chain of events is set into play. Sound vibrates in the air and a wave travels to my ear. The wave is processed and analyzed into frequencies inside the ear, and a signal is sent to the auditory cortex. Further processing takes place here: isolations of certain aspects of the signal, categorization, and ultimately reaction. All this is not so hard to understand in principle. But why should this be accompanied by an *experience*? And why, in particular, should it be accompanied by *that* experience, with its characteristic rich tone and timbre?”

3. The argument is characterized by *methodological panpsychism*. Panpsychism is used methodologically in the sense that it is discharged in the third premiss. The relevant point of methodological panpsychism is to deny that there are qualia inside biological or non-biological brains *only*. That is, panpsychism here is understood in a *minimalist* sense: it is *at least* the thesis that there are qualia inside *and* outside brains. ‘Panpsychism’ is a bit of a misnomer - a more appropriate term would be ‘panqualitativism’ – but I’ll keep the label in this paper.

4. Physicalism as used in the third premiss is *relevant* physicalism. It is not the thesis that everything is physical. Maybe there are non-physical facts; maybe, say, facts about beliefs are not physical facts. But the reality that is relevant to qualia is physical reality. Relevant physicalism denies that there are, *on top* of physical facts, phenomenal facts.

5. How does panpsychism solve the hard problem of consciousness? According to traditional physicalism, qualia are not just “correlated” with any specifically brain event or property or configuration. They *are* the occurrence of that very configuration, the instantiation of a physical property. That much is endorsed by methodological panpsychism as well. Traditional physicalism suggests that we look at which properties are qualia, in the wide catalogue of physical properties. Methodological panpsychism should be best conceived as an *extension* of physicalism. The physicalist admits that certain physical states are qualitative states. The methodological panpsychist extends this conception to all physical states. If there are qualia independently of the existence of brains, then there is no problem of explaining how the brain, or particular classes of brain events, can *produce* or *realize* qualitative states. (The correct picture, as we shall see, should be that the brain *uses* qualia that are *already* available in nature.)

6. Panpsychism relieves us from the problem of understanding the seeming exceptionality of our conscious life in the natural world. But in order to properly understand panpsychism, we need to explore two things. If qualia exist outside brains, then they exist, say, in this stone in front of me. What does this mean? The hard problem of consciousness was one of making sense of how qualia in my brain can arise from physical facts only, and now

we seem to be faced with the dual problem of making sense of how a stone can be inhabited by qualia. There is a second problem. If the qualia in the stone are, at bottom, the same type of phenomenon as the qualia in my brain, how can we make sense of the intuition that what goes on inside the stone is after all quite different from what goes on inside my brain? How can we deny that the stone is conscious, and what would that mean?

7. Premiss (3) discharges the psychistic overtones of panpsychism. The stone is inhabited by qualia because physical reality is qualitative. Physical reality *is* just that - qualitative reality. The stone is completely described by physics; there are no non-physical states in the stone. And there are no non-physical states in my brain. By understanding that the stone is pure physics, we should be facilitated in seeing how our own qualia are pure physics. Qualia patterns in my head *are* just a portion of the physical world. There may be *more* to the physical world, but there is *at least* this.

8. Panspsychism is entailed by physicalism. There aren't two sides, so to speak, of reality in the stone (or in my brain) - a physical side, and a psychical side, that are magically brought together by the third premiss. There is no *further* "qualitative aspect" of physical reality.

It is worth considering four theses that come quite close to (3) without coinciding with it:

- (a) Traditional physicalist theories of qualia are different from (3) because they allow only for a subclass of brain events to be candidates for conscious events.
- (b) Neutral monism is different from (3) because it postulates a reality which in itself is neither physical nor qualitative, and can manifest itself as either physical or qualitative in different contexts. This is almost the correct view, except that it incorrectly denies qualia to the stone, and incorrectly denies physical reality to qualia in my brain.
- (c) The postulation of Leibnitian monads is different from (3) because monads account for the difference between me and the stone in terms of *degrees* of consciousness.
- (d) The thesis that consciousness is information (Chalmers 1996) is different from (3) because it identifies qualitative states with information (or with causal transactions), i.e. with a subclass of physical facts.

9. Traditional physicalist theories deny qualia to the stone, hence they are incompatible with panpsychism. On the other hand, there is a common confusion behind both neutral monism and Leibnitian monads, and it is a confusion between phenomenal consciousness and representation. For I have qualia exactly as the stone has it. But my qualia are related to me and to the world by a complex representational (intentional) network, whereas the qualia that occur inside the stone are not so related to the stone and to the world. Here lies the difference: My body is the result of a complex evolutionary process that has *domesticated* existing qualia (physical properties, once more: as per premisses (2) and (3)) so as to allow me to use them for representational purposes. Representational networks completely account for the differences between me and the stone. Representational networks also account for differences between my present consciousness and subpersonal types of 'consciousness' (in the loose sense). There is again a confusion between consciousness and representation. Subpersonal states may be representational. Their difference from personal states is in the way they are related to my body and to other representational states. But there is no reason whatsoever to deny qualitiveness to subpersonal states. Only, they are qualia that I do not use.

A good phenomenological example would be absent-minded driving. It looks as if the consensus is that it is a form of unconscious activity, in the sense that it is performed subpersonally, with no access by consciousness. Is the qualitative aspect suspended as well? I find this phenomenologically implausible. When I emerge from absent-minded driving it does not feel like I am waking up. I know that I was not zombied in the minutes before surfacing. Which would mean that I can accept, even at the level of common sense, the idea that qualia are there that are not accessed by consciousness.

10. What is the advantage to be gained from the physicalism of (3) over other, less liberal forms of physicalism? And what right do we have to assert (3)? The basic idea is that physicalism as we know it is unbelievably restrictive relative to the vast wealth of existing physical properties. This restrictive conception has two consequences. First, it is difficult to understand at all traditional physicalism as a way of solving the hard problem of consciousness. Second, it is mandatory to accept as a miracle that the physical properties identified by traditional physicalism with qualia have been selected for representational purposes in the course of evolution. Building on these two intuitions, two arguments could be invoked here in order to establish premiss (3), the *Argument from symmetry* and the *argument from evolutionary blindness*. The two arguments are correlated.

The argument from symmetry. According to traditional physicalism, *all* qualia just are identical to physical properties. This is tantamount to saying that *some* physical properties are qualitative. I insist on this way of reading the identity claim of physicalism. Typically, one focuses on the attempt to understand the identity

statement in the left-to-right sense (the sense in which one reads it: identities themselves have no direction): How can this particular feeling of pain be, at bottom, just a physical property? But it is much more interesting to focus on the backwards reading. We read that some physical facts are just qualitative facts. It says that there is a quality to at least some physics.

Identities do not need explanations (they are what does the explaining), but still something prevents us from getting a satisfactory picture of how qualia fit in the physical world. Suppose phenomenal consciousness is some event that has to do with a 40Hz cycle of electric discharges – as per some actual proposal. Is a 41Hz cycle not qualitative? Why? What on earth makes a 41Hz shot a dark, dumb, “purely” physical event, and 40Hz a qualia feast? Is 40Hz a specific constant of the universe?

If you are a physicalist, you are bound to accept that some physical facts are qualitative facts (per the physicalist identities). But then the hard problem of consciousness resurfaces in a new way. Why is physical property P a qualitative property and not some other property P' that only slightly (or wildly, for that matter) differs from P? Why is P a quale, and P' a zombie property, as we may call it?

The challenge for the traditional physicalist is to create an antipanpsychist confinement of phenomenal consciousness to some physical properties only. It has to be a confinement which does not reinstate the hard problem of consciousness in a new disguise.

The argument from evolutionary blindness. Moreover, how could evolution have selected it P and not P'? Natural selection is blind to physical structure. A certain property P, we are told, is instantiated by a certain portion of the brain. But now assume that P is the only physical property that is qualitative. Evolution cannot discover isolated spikes in a fitness landscape. It follows that had evolution picked up a slightly different property we would have been zombies¹.

In a nutshell. Traditional physicalism is unable to draw an useful distinction between conscious physical properties and zombie properties. It is unable to tell us by which evolutionary subtlety we did not turn out to be zombies. It makes evolution much too far-sighted and miraculous.

The entailment of panpsychism by physicalism finishes the physicalistic job. What is the result? All physical properties are qualitative, but in a trivial way. It is not necessary to postulate that physical properties are endowed with special epiphenomenal quality. Being qualitative is just their mode of being. This is what (instantiated) physical reality is – qualitative reality. At this stage we can rehearse the principle that the force of identities is that they require no explanations.

I can offer an analogy here. It may be argued that there exists a difficult explanatory gap to be filled when one tries to understand what it is for a property P to be instantiated by an object. You may understand perfectly what mass is, and still fail to understand what it is for it to be instantiated in a given body. Something seems to have been added to the property. A physical theory does not deliver any explanation here. Instantiation is the mode of being of properties. But, then, there really is nothing to explain here. There are no metaphysical explanations.

This closes the metaphysical problem of qualia. There remain some nontrivial, nonmetaphysical problems.

11. Qualia are the way of being of instantiated physical reality; they are not an added property thereof. They are just what physical reality is. Qualia have many forms, some of which we are in when we use them - green “flashes” in my mind, next to yellow flashes. These flashes are the way reality is (plausibly at some micro level). Now, the scheme of things is arguably larger. Consciousness is usually described as consciousness *for* me, and *of* some external object. How are the *for* and the *of* to be understood? These are non trivial, nonmetaphysical problems for a theory of consciousness. A naturalistic picture goes as follows.

(a) The *of* problem is the problem of representation. Close your eyes. Phosphenes pop out in the dark. You cannot help but use them to inform you of the presence, somewhere in front of you, of moving coloured luminescent patches. You are wrong, of course, but this is a price to pay for the success of the normal workings of perception. Open your eyes and move your hand in front of you. Qualia pop out in your mind in a way that covaries very tightly with the movement of your hand. Evolution has solved the problem of representation in a stable way.

¹ Strictly speaking, there could be other unexplored possibilities here. There may be various types of ‘being qualitative’ (not in the sense of the determinate quale, such as the red or the itchy quale, but in the sense of the determinable ‘qualitative’). In such a case an evolutionary uphill climbing strategy may be possible. However, the idea of different determinable qualitativenesses strikes me as a wild hypothesis.

Some types of qualia (or of patterns of qualia) have been recruited and domesticated by living beings in the course of evolution for representational purposes of various sorts. So well are they domesticated for those purposes that it has become impossible for us to use those instances and not have the impression of representing something.

(b) The *for* problem is the problem of the difference between personal and subpersonal states. Some types of qualia patterns are recruited and domesticated by living beings as *available* for them for representational purposes. So well they have been domesticated that it has become impossible for us to represent something and not have the impression of *using* them. Our relation to these qualia is what explains the impression that they are our qualia.

Note that the thesis that there are qualia in the stone does not entail that the stone is or can be conscious. The stone has no ancestors that domesticated qualia for representing the world to them. Or, if you prefer, the stone is not a subject.

12. The correctness (a normative notion) of the pattern of qualia variation is explained, I assume, by the theory of evolution. I assume that representational networks for qualia are completely naturalized. This is not to claim that the whole of what is ordinarily taken as the 'mental' is naturalized or even naturalizable in principle. Beliefs may be not naturalizable. This fact would not be a counterexample to the methodological panpsychist solution to the hard problem, though, because there is no "feel" associated to beliefs, no way it is like to believe that giraffes are tall. Beliefs are "conscious" in the loose and popular sense - but this may be different from the sense in which there is consciousness around when I am related to a quale. This caveat allows us to select the relevant form of physicalism in premiss (2); we may want physicalism to be weaker than the thesis that everything is physical.

The differences between the physical world and the mental world are not a counterexample to thesis (2), nor are they a problem for panpsychism. Actually, I would consider an advantage of the position put forward here that the so called "mind-body problem" becomes a mixed bag of quite different problems (qualia, normativity, subjectivity, etc.), each with a different solution or dissolution.

13. The existence of qualitatively endowed robots or electronic hardware of various sorts is compatible with the present solution. As qualia manifest themselves in stones, so do they in robots (pieces of physical reality). There may be quite important differences though. Robots may not be using their qualia in a proper representational way. They may be instantiating representational states, but still be quite different from us because their qualia are not used by them. A robot endowed with qualia may be such as to appear to us *not* to pass the Chinese Room test. The Chinese Room experiment tests the attribution of intentionality but it does not test the possibility of attributing qualia.

14. We said that the qualitative aspect is just physics, but can we understand more substantially what qualia are? What is a given quale? An answer comes from the fact that qualia are used in representation, and quite successfully so. We can hope to discover some of the relevant dimensions of qualia by studying some of the structural dimensions of representation. Representation imposes *some* constraints on the physics that is used for the representing.

This heuristic is not infallible. First of all, there may be qualia whose existence and whose type we are never aware of. (This thesis of the transcendence of qualitiveness is just a logical consequence of the conjunction of realism about physical reality and of panpsychism.) This may be the case because a certain type of qualia is unsuitable for being used for representation, or it has never been used for representation because of preemptive domestication. Some qualia have been used to represent colours, and their availability as representations may have preempted the use of other, maybe more appropriate qualia.

For the sake of the argument, one can make two assumptions about visual consciousness. The first (Jackendoff 1987, p. 293) is that the content of visual awareness is a 2.5-D sketch, a distribution of colours at some distance and in some direction from an observer. (The thesis entails that we are not visually aware of there being dogs in front of us, and this requires an independent discussion.) The second assumption is that some (micro)physical reality has an appropriate number of dimensions in order to be used as a (part of a) representation that has the format of the 2.5-D sketch. In our case, qualia would need to have five dimensions, for hue, darkness, saturation, direction and distance. This is where we should look for in the world in order to spell out a substantial theory of qualia.

The dimensions of qualitiveness ought to be there *before* any living being or any mechanism that uses them. This excludes that qualia could *result* from the functioning of living beings and mechanisms (and is hence incompatible with informational versions of panpsychism.)

15. The argument *forces* a solution on the hard problem of consciousness by denying that physics is non-phenomenal. 'Forcing' means to accept the truth of a questionable premiss; it delivers the solution, at a cost. My purpose is to show that once the premiss is properly understood, costs evaporate. The advantage of forcing is that it prevents objections deriving from misconceptions that have no reasons to be there once the solution is established. This may be methodologically appropriate if we are after *deep* explanations and theories, not just reconceptualizations or redescriptions of conscious life, although reconceptualizations may help us in understanding the nature of consciousness.

16. To sum up. Methodological panpsychism assumes the following: the truth of relevant physicalism; a minimalist form of panpsychism, according to which some qualia occur outside brains; the (possibly forthcoming) truth of naturalistic accounts of perception and of the distinction between personal and subpersonal states. The minimalist form of panpsychism becomes innocent once we accept physicalism. The argument solves the hard problem of consciousness by dissolving it. There is no problem of "finding" the "physical correlate" of qualia and then to understand "how" that physical correlate "could be" qualitative. The *mode of being* of physical reality is assumed to be qualitative from the onset. But it is an innocent way of being qualitative. Actually, upon reflection, we discover that we do not have a single argument that could make us think that physical reality it is not qualitative! We do have thought experiments that presuppose a qualitativeless conception of physics and hence cannot prove it. Intuitions to the contrary (the "Cartesian intuition", the idea that we would never understand what it is like to be a bat, etc.) are not sufficient to establish the non-qualitativeness of physical reality. These intuitions actually have a simple explanation in the fact that we pre-philosophically spend a good deal of energy in attributing psychological states of various kind to people and animals in order to explain their behaviour, and do not do the same with, say, stones. Conscious states do not play any role in explanation of the behaviour of stones. It is tempting to conclude that this difference is explained by the fact that stones are just inert matter; we are matter full of qualia – as we can experience all the time. But the conclusion is unwarranted, and have seen that it is easily overruled by other ways to account for the difference between the stone and us.

And why do we have hard times in understanding the other thesis, that qualitative reality is physical? Again, the dual "Cartesian intuition" is at the origin of the bias: the idea that physical reality is the domain of *res extensa*, of the physical bodies. An object bias pervades our representation of the world and an intentional bias pervades our representation of the mind..

The universe inside the stone is somewhat different from ours, but not much. In both there is a "blooming, buzzing" reality, in William James's phrase. But inside the stone there is a "blooming, buzzing confusion", whereas inside us there is an ordered show that our bodies can use to navigate their environment. We (our biological ancestors, that is) have domesticated qualia, that is to say, we have domesticated the physical world.

I am indebted to Achille Varzi, Kevin Mulligan, Gianfranco Soldati, Diego Marconi, Nirmalangshu Mukherji, Michele Di Francesco, Alberto Voltolini, Pascal Ludwig, and the audiences at Institut Nicod, Paris, at the University of Vercelli, at Jadavpur University in Calcutta and at Indian Council of Philosophical Research in Delhi for helpful discussions.

Chalmers, D., 1996, *The Conscious Mind*, Oxford: Oxford University Press.

Jackendoff, R., 1987, *Consciousness and the Computational Mind*, Cambridge, Mass., MIT Press.

