Journal of Consciousness Studies, **3**, No. 5–6, 1996, pp. 402–8

THE HORNSWOGGLE PROBLEM[1]

Patricia Smith Churchland, Department of Philosophy,
University of California at San Diego, La Jolla, CA 92093, USA.

Abstract: Beginning with Thomas Nagel, various philosophers have propsed setting conscious experience apart from all other problems of the mind as 'the most difficult problem'. When critically examined, the basis for this proposal reveals itself to be unconvincing and counter-productive. Use of our current ignorance as a premise to determine what we can never discover is one common logical flaw. Use of 'I-cannot-imagine' arguments is a related flaw. When not much is known about a domain of phenomena, our inability to imagine a mechanism is a rather uninteresting psychological fact about us, not an interesting metaphysical fact about the world. Rather than worrying too much about the meta-problem of whether or not consciousness is uniquely hard, I propose we get on with the task of seeing how far we get when we address neurobiologically the problems of mental phenomena.

I: Introduction

Conceptualizing a problem so we can ask the right questions and design revealing experiments is crucial to discovering a satisfactory solution to the problem. Asking where animal spirits are concocted, for example, turns out not to be the right question to ask about the heart. When Harvey asked instead, 'How much blood does the heart pump in an hour?', he conceptualized the problem of heart function very differently. The reconceptualization was pivotal in coming to understand that the heart is really a pump for circulating blood; there are no animal spirits to concoct. My strategy here, therefore, is to take the label, 'The Hard Problem' (Chalmers, 1995) in a constructive spirit — as an attempt to provide a useful conceptualization concerning the very nature of consciousness that could help steer us in the direction of a solution. My remarks will focus mainly on whether in fact anything positive is to be gained from the 'hard problem' characterization, or whether that conceptualization is counterproductive.

I cannot hope to do full justice to the task in short compass, especially as this characterization of the problem of consciousness has a rather large literature surrounding it. The watershed articulation of this view of the problem is Thomas Nagel's classic paper 'What is it like to be a bat?' (1974) In his opening remarks, Nagel comes straight to the point: 'Consciousness is what makes the mind–body problem really intractable.' Delineating a contrast between the problem of consciousness and all other mind–body problems, Nagel asserts: 'While an account of the physical basis of mind must explain many things, this [conscious experience] appears to be the most difficult.' Following Nagel's lead, many other philosophers, including Frank Jackson, Saul Kripke, Colin McGinn, John Searle, and most recently, David Chalmers, have extended and developed Nagel's basic idea that consciousness is not tractable neuroscientifically.

Although I agree that consciousness is, certainly, *a* difficult problem, difficulty *per se* does not distinguish it from oodles of other neuroscientific problems. Such as how the brains of homeotherms keep a constant internal temperature despite varying external conditions. Such as the brain basis for schizophrenia and autism. Such as why we dream and sleep. Supposedly, something sets consciousness apart from *all* other macro-function

[1] This paper is based on a talk I presented at the 'Tuscon II' conference, 'Toward a Science of Consciousness' held at Tucson, Arizona, in April 1996. Many thanks are owed to the organizers of the meeting, and thanks also to Paul Churchland, David Rosenthal, Rodolfo Llinás, Michael Stack, Dan Dennett, Ilya Farber and Joe Ramsay for advice and ideas.

brain riddles such that it stands alone as The Hard Problem. As I have tried to probe precisely what that is, I find my reservations multiplying.

II: Carving Up the Problem Space

The-Hard-Problem label invites us to adopt a principled empirical division between consciousness (the hard problem) and problems on the 'easy' (or perhaps hard but not Hard?) side of the ledger. The latter presumably encompass problems such as the nature of short-term memory, long-term memory, autobiographical memory, the nature of representation, the nature of sensori-motor integration, top-down effects in perception — not to mention such capacities as attention, depth perception, intelligent eye movement, skill acquisition, planning, decision-making, and so forth. On the other side of the ledger, all on its own, stands consciousness — a uniquely hard problem.

My lead-off reservation arises from this question: what is the rationale for drawing the division exactly there? Dividing off consciousness from all of the so-called 'easy problems' listed above implies that we could understand all those phenomena and still not know *what it was for* . . . what? The 'qualia-light' to go on? ? Is *that* an insightful conceptualization? What exactly is the evidence that we could explain all the 'easy' phenomena and still not understand the neural mechanisms for consciousness? (Call this the 'left-out' hypothesis.) That someone can *imagine* the possibility is not *evidence* for the real possibility. It is only evidence that somebody or other *believes* it to be a possibility. That, on its own, is not especially interesting. Imaginary evidence, needless to say, is not as interesting as real evidence, and what needs to be produced is some real evidence.[2]

The left-out hypothesis — that consciousness would still be a mystery, even if we could explain all the easy problems — is dubious on another count: it begs the question against those theories that are exploring the possibility that functions such as attention and short-term memory are crucial elements in the consciousness (see especially Crick, 1994; P.M. Churchland, 1995). The rationale sustaining this approach stems from observations such as that awake persons can be unaware of stimuli to which they are not paying attention, but can become aware those stimuli when attention shifts. There is a vast psychological literature, and a nontrivial neuroscientific literature, on this topic. Some of it powerfully suggests that attention and awareness are pretty closely connected. The approach might of course be wrong, for it is an empirical conjecture. But if it is wrong, it is wrong because of the *facts*, not because of an arm-chair definition. The trouble with the 'hard problem' characterization is that *on the strength of a proprietary definition*, it rejects them as wrong. I do find that unappealing, since the nature of consciousness is an empirical problem, not problem that can be untangled by semantic gerrymandering.

[2] As I lacked time in my talk at Tucson to address the 'Mary' problem, a problem first formulated by Frank Jackson in 1982, let me make several brief remarks about it here. In sum, Jackson's idea was that there could exist someone, call her Mary, who knew everything there was to know about how the brain works but still did not know what it was to see the colour green (suppose she lacked 'green cones' , to put it crudely). This possibility Jackson took to show that qualia are therefore not explainable by science. The main problem with the argument is that to experience green qualia, certain wiring has to be in place in Mary's brain, and certain patterns of activity have to obtain and since, by Jackson's own hypothesis, she does not have that wiring, then presumably the relevant activity patterns in visual cortex are not caused and she does not experience green. Who would expect her visual cortex — V4, say — would be set ahumming just by virtue of her *propositional* (linguistic) knowledge about activity patterns in V4? Not me, anyhow. She can have propositional knowledge via other channels, of course, including the knowledge of what her own brain lacks *vis à vis* green qualia. Nothing whatever follows about whether science can or cannot explain qualia.

What drives the left-out hypothesis? Essentially, a thought-experiment, which roughly goes as follows: we can conceive of a person, like us in all the aforementioned easy-to-explain capacities (attention, short term memory, etc.), but lacking qualia. This person would be *exactly* like us, save that he would be a Zombie — an anaqualiac, one might say. Since the scenario is conceivable, it is possible, and since it is possible, then whatever consciousness is, it is explanatorily independent of those activities.[3]

I take this argument to be a demonstration of the feebleness of thought-experiments. *Saying* something is possible does not thereby guarantee it *is* a possibility, so how do we know the anaqualiac idea is really possible? To insist that it must be is simply to beg the question at issue. As Francis Crick has observed, it might be like saying that one can imagine a possible world where gases do not get hot, even though their constituent molecules are moving at high velocity. As an argument against the empirical identification of temperature with mean molecular KE, the thermodynamic thought-experiment is feebleness itself.

Is the problem on the 'hard' side of the ledger sufficiently well-defined to sustain the division as a fundamental empirical principle? Although it is easy enough to agree about the presence of qualia in certain prototypical cases, such as the pain felt after a brick has fallen on a bare foot, or the blueness of the sky on a sunny summer afternoon, things are less clear-cut once we move beyond the favoured prototypes. Some of our perceptual capacities are rather subtle, as, for example, positional sense is often claimed to be. Some philosophers, e.g. Elizabeth Anscombe, have actually opined that we can know the position of our limbs without any 'limb-position' qualia. As for me, I am inclined to say I do have qualitative experiences of where my limbs are — it feels different to have my fingers clenched than unclenched, even when they are not visible. The disagreement itself, however, betokens the lack of consensus once cases are at some remove from the central prototypes.

Vestibular system qualia are yet another non-prototypical case. Is there something 'vestibular-y' it feels like to have my head moving? To know which way is up? Whatever the answer here, at least the answer is not glaringly obvious. Do eye movements have eye-movement qualia? Some maybe do, and some maybe do not. Are there 'introspective qualia', or is introspection just paying attention to perceptual qualia and talking to yourself? Ditto, plus or minus a bit, for self-awareness. Thoughts are also a bit problematic in the qualia department. Some of my thoughts seem to me to be a bit like talking to myself and hence like auditory imagery but some just come out of my mouth as I am talking to someone or affect decisions without ever surfacing as a bit of inner dialogue. None of this is to deny the pizzazz of qualia in the prototypical cases. Rather, the point is just that prototypical cases give us only a *starting point* for further investigation, and nothing like a full characterization of the class to which they belong.

My suspicion with respect to The Hard Problem strategy is that it seems to take the class of conscious experiences to be much better defined than it is. The point is, if you are careful to restrict your focus to the prototypical cases, you can easily be hornswoggled into assuming the class is well-defined. As soon as you broaden your horizons, troublesome questions about fuzzy boundaries, about the connections between attention, short term memory and awareness, are present in full, what-do-we-do-with-*that* glory.

Are the easy problems known to be easier than The Hard Problem? Is the hard/easy division grounded in fact? To begin with, it is important to acknowledge that for none of the so-called 'easy' problems, do we have an understanding of their solution (see the partial list on p. 403). It is just false that we have anything approximating a comprehen-

[3] Something akin to this was argued by Saul Kripke in the 1970's.

sive theory of sensori-motor control or attention or short-term memory or long-term memory. Consider one example. A signature is recognizably the same whether signed with the domi- nant or non-dominant hand, with the foot, with the mouth or with the pen strapped to the shoulder. How is 'my signature' represented in the nervous system? How can completely different muscle sets be invoked to do the task, even when the skill was not acquired using those muscles? We do not understand the general nature of motor representation.

Notice that it is not merely that we are lacking details, albeit important details. The fact is, we are lacking important conceptual/theoretical ideas about how the nervous system performs fundamental functions — such as time management, such as motor control, such as learning, such as information retrieval. We do not understand the role of back projections, or the degree to which processing is organized hierarchically. These are genuine puzzles, and it is unwise to 'molehill' them in order to 'mountain' up the problem of consciousness. Although quite a lot is known at the cellular level, the fact remains that how real neural networks work and how their output properties depend on cellular properties still abounds with nontrivial mysteries. Naturally I do not wish to minimize the progress that has been made in neuroscience, but it is prudent to have a cautious assessment of what we really do not yet understand.

Carving the explanatory space of mind–brain phenomena along the hard and the easy line, as Chalmers proposes, poses the danger of inventing an explanatory chasm where there really exists just a broad field of ignorance. It reminds me of the division, deep to medieval physicists, between sublunary physics (motion of things below the level of the moon) and superlunary physics (motion of things above the level of the moon). The conviction was that sublunary physics was tractable, and it is essentially based on Aristotelian physics. Heavy things fall because they have gravity, and fall to their natural place, namely the earth, which is the centre of the universe. Things like smoke have levity, and consequently they rise, *up* being their natural place. Everything in the sublunary realm has a 'natural place', and that is the key to explaining the behaviour of sublunary objects. Superlunary events, by contrast, we can neither explain nor understand, but in any case, they have neither the gravity nor levity typical of sublunary things.

This old division was not without merit, and it did entail that events such as planetary motion and meteors were considered unexplainable in terrestrial terms, but probably were divinely governed. Although I do not know that Chalmers' easy/hard distinction will prove ultimately as misdirected as the sublunary/superlunary distinction, neither do I know it is any more sound. What I do suspect, however, is that it is much too early in the science of nervous systems to command much credence.

The danger inherent in embracing the distinction as a principled empirical distinction is that it provokes the intuition that only a real humdinger of a solution will suit The Hard Problem. Thus the idea seems to go as follows: the answer, if it comes at all, is going to have to come from somewhere Really Deep — like quantum mechanics — or perhaps it requires a whole new physics. As the lone enigma, consciousness surely cannot be just a matter of a complex dynamical system doing its thing. Yes, there are emergent properties from nervous systems such as co-ordinated movement as when an owl catches a mouse, but consciousness (the hard problem) is an emergent property like unto no other. Consequently, it will require a very deep, very radical solution. That much is evident sheerly from the hardness of The Hard Problem.

I confess I cannot actually see that. I do not know anything like enough to see how to solve either the problem of sensori-motor control or the problem of consciousness. I

certainly cannot see enough to know that one problem will, and the other will not, require a Humdinger solution.

III: Using Ignorance as a Premise

In general, what substantive conclusions can be drawn when science has not advanced very far on a problem? Not much. One of the basic skills we teach our philosophy students is how to recognize and diagnose the range of nonformal fallacies that can undermine an ostensibly appealing argument: what it is to beg the question, what a *non sequitur* is, and so on. A prominent item in the fallacy roster is *argumentum ad ignorantiam* — argument from ignorance. The canonical version of this fallacy uses ignorance as the key premise from which a substantive conclusion is drawn. The canonical version looks like this:

> We really do not understand much about a phenomenon *P*. (Science is largely ignorant about the nature of *P*.)
>
> Therefore: we *do* know that:
>
> (1) *P* can never be explained, or
> (2) Nothing science could ever discover would deepen our understanding of *P*, or
> (3) *P* can never be explained in terms of properties of kind *S*.

In its canonical version, the argument is obviously a fallacy: none of the tendered conclusions follow, not even a little bit. Surrounded with rhetorical flourish, much brow furrowing and hand-wringing, however, versions of this argument can hornswoggle the unwary.

From the fact that we do not know something, nothing very interesting follows — we just don't know. Nevertheless, the temptation to suspect that our ignorance is telling us something positive, something deep, something metaphysical or even radical, is ever-present. Perhaps we like to put our ignorance in a positive light, supposing that but for the Profundity of the phenomenon, we *would* have knowledge. But there are many reasons for not knowing, and the specialness of the phenomenon is, quite regularly, not the real reason. I am currently ignorant of what caused an unusual rapping noise in the woods last night. Can I conclude it must be something special, something unimaginable, something . . . alien . . . other-worldly? Evidently not. For all I can tell now, it might merely have been a raccoon gnawing on the compost bin. Lack of evidence for something is just that: lack of evidence. It is not positive evidence for something else, let alone something of a humdingerish sort. That conclusion is not very glamorous perhaps, but when ignorance is a premise, that is about all you can grind out of it.

Now if neuroscience had progressed as far on the problems of brain function as molecular biology has progressed on transmission of hereditary traits, then of course we would be in a different position. But it has not. The only thing you can conclude from the fact that attention is mysterious, or sensori-motor integration is mysterious, or that consciousness is mysterious, is that we do not understand the mechanisms.

Moreover, the mysteriousness of a problem is not a fact about the problem, it is not a metaphysical feature of the universe — it is an epistemological fact about *us*. It is about where we are in current science, it is about what we can and cannot understand, it is about what, given the rest of our understanding, we can and cannot imagine. It is not a property of the problem itself.

It is sometimes assumed that there can be a valid transition from 'we cannot now explain' to 'we can never explain' , so long as we have the help of a subsidiary premise,

namely, 'I cannot *imagine* how we could ever explain . . .' But it does *not* help, and this transition remains a straight-up application of argument from ignorance. Adding 'I cannot imagine explaining *P*' merely adds a psychological fact about the speaker, from which again, nothing significant follows about the nature of the phenomenon in question. Whether we can or cannot imagine a phenomenon being explained in a certain way is a psychological fact about us, not an objective fact about the nature of the phenomenon itself. To repeat: it is an epistemological fact about what, given our current knowledge, we can and cannot understand. It is not a metaphysical fact about the nature of the reality of the universe.

Typical of vitalists generally, my high school biology teacher argued for vitalism thus: I cannot *imagine* how you could get living things out of dead molecules. Out of bits of proteins, fats, sugars — how could life itself emerge? He thought it was obvious from the sheer mysteriousness of the matter that it could have no solution in biology or chemistry. He assumed he could tell that it would require a Humdinger solution. Typical of lone survivors, a passenger of a crashed plane will say: I cannot imagine how I alone could have survived the crash, when all other passengers died instantly. Therefore God must have plucked me from the jaws of death.

Given that neuroscience is still very much in its early stages, it is actually not a very interesting fact that someone or other cannot imagine a certain kind of explanation of some brain phenomenon. Aristotle could not imagine how a complex organism could come from a fertilized egg. That of course was a fact about Aristotle, not a fact about embryogenesis. Given the early days of science (500 BC), it is no surprise that he could not imagine what it took many scientists hundreds of years to discover. I cannot imagine how ravens can solve a multi-step problem in one trial, or how temporal integration is achieved, or how thermoregulation is managed. But this is a (*not very interesting*) psychological fact about me. One could, of course, use various rhetorical devices to make it seem like an interesting fact about me, perhaps by emphasizing that it is a really really hard problem; but if we are going to be sensible about this, it is clear that my inability to imagine how thermoregulation works is *au fond*, pretty boring.

The 'I-cannot-imagine' gambit suffers in another way. Being able to imagine an explanation for *P* is a highly open-ended and under-specified business. Given the poverty of delimiting conditions of the operation, you can pretty much rig the conclusion to go whichever way your heart desires. Logically, however, that flexibility is the kiss of death.

Suppose someone claims that she *can* imagine the mechanisms for sensori-motor integration in the human brain but *cannot* imagine the mechanisms for consciousness. What exactly does this difference amount to? Can she imagine the former in *detail*? No, because the details are not known. What is it, precisely, that she can imagine? Suppose she answers that in a very general way she imagines that sensory neurons interact with interneurons that interact with motor neurons, and via these interactions, sensori-motor integration is achieved. Now if that is all 'being able to imagine' takes, one might as well say one can imagine the mechanisms underlying consciousness. Thus: 'The interneurons do it.' The point is this: if you want to contrast being able to imagine brain mechanisms for attention, short term memory, planning etc., with being unable to imagine mechanisms for consciousness, you have to do more that say you can imagine neurons doing one but cannot imagine neurons doing the other. Otherwise one simply begs the question.

To fill out the point, consider several telling examples from the history of science. Before the turn of the twentieth century, people thought that the problem of the precession of the perihelion of Mercury was essentially trivial. It was annoying, but ultimately, it would sort itself out as more data came in. With the advantage of hindsight, we can see

that assessing this as an easy problem was quite wrong — it took the Einsteinian revolution in physics to solve the problem of the precession of the perihelion of Mercury. By contrast, a really hard problem was thought to be the composition of the stars. How could a sample ever be obtained? With the advent of spectral analysis, that turned out to be a readily solvable problem. When heated, the elements turn out to have a kind of fingerprint, easily seen when light emitted from a source is passed through a prism.

Consider now a biological example. Before 1953, many people believed, on rather good grounds actually, that in order to address the copying problem (transmission of traits from parents to offspring), you would first have to solve the problem of how proteins fold. The former was deemed a much harder problem than the latter, and many scientists believed it was foolhardy to attack the copying problem directly. As we all know now, the basic answer to the copying problem lay in the base-pairing of DNA, and it was solved first. Humbling it is to realize that the problem of protein folding (secondary and tertiary) is *still* not solved. *That*, given the lot we now know, does seem to be a hard problem.

What is the point of these stories? They reinforce the message of the argument from ignorance: from the vantage point of ignorance, it is often very difficult to tell which problem is harder, which will fall first, what problem will turn out to be more tractable than some other. Consequently our judgments about relative difficulty or ultimate tractability should be appropriately qualified and tentative. Guesswork has a useful place, of course, but let's distinguish between blind guesswork and educated guesswork, and between guesswork and confirmed fact. The philosophical lesson I learned from my biology teacher is this: when not much is known about a topic, don't take terribly seriously someone else's heartfelt conviction about what problems are scientifically tractable. Learn the science, do the science, and see what happens.

References

Chalmers, David J. (1995), 'Facing up to the problem of consciousness', *Journal of Consciousness Studies*, **2** (3), pp. 200–19.

Churchland, Paul M. (1995), *The Engine of Reason; The Seat of the Soul* (Cambridge, MA: MIT Press).

Crick, Francis (1994), *The Astonishing Hypothesis* (New York: Scribner and Sons).

Jackson, Frank (1982), 'Epiphenomenal qualia', *Philosophical Quarterly*, **32**, pp. 127–36.

Nagel, Thomas (1974), 'What is it like to be a bat?', *Philosophical Review*, **83**, pp.435–50.