

Is the Hard Problem of Consciousness Universal?*

David J. Chalmers

The hard problem of consciousness is the problem of explaining how and why physical processes give rise to consciousness. The meta-problem of consciousness is the problem of explaining why there seems to be a hard problem of consciousness. Equivalently, it is the problem of explaining why people have problem intuitions: dispositions to make certain key judgments that underlie the problem of consciousness.

The problem intuitions in question include metaphysical intuitions (“consciousness is non-physical”), explanatory intuitions (“physical processes can’t fully explain consciousness”), knowledge intuitions (“someone who knows all about the brain but has never seen colors doesn’t know what it’s like to see red”), and modal intuitions (“we can imagine all these physical processes without consciousness”). There are also intuitions about the value of consciousness, the distribution of consciousness, and more.

One aspect of the meta-problem is the question: is the hard problem of consciousness universal? In its starkest form, one could put this as a question about *judgment universality*: does everyone (or at least every normal adult human) judge that there is a hard problem of consciousness? Does everyone make judgments along the lines of the intuitions above? The answer to this question is obviously no. Some theorists reject these judgments. Many people never consider them. In some cultures, the issue never seems to have been raised.

Still, there are weaker theses nearby that cannot be refuted so easily. We could ask a question about *intuition universality*: are some problem intuitions universal? That is, does everyone at least have some disposition or tendency to make certain judgments that underlie the hard problem? The existence of theorists who reject the problem is consistent with this claim, as long as the theorists have a tendency that is overridden by some other factor. The presence of people who

⁰This article is one of three responding to 39 commentaries in a symposium on my article “The Meta-Problem of Consciousness” in the *Journal of Consciousness Studies*. I’m grateful to all the commentators for their interesting contributions, and to François Kammerer for his editorial work and feedback.

never consider the problem is also consistent with this claim, as long as they have a disposition or tendency to make the relevant judgments if they consider the issue.

My sense is that at least for some problem intuitions, a disposition or tendency of this sort is present in a good majority of people, at least among reflective people in contemporary Western cultures. Even people who reject the hard problem often acknowledge some intuitions in its favor. At the same time, it is far from obvious that such a disposition or tendency is present in everyone. There are at least some people who appear to have very little tendency to accept any of the problem intuitions. So considered as a thesis about all individuals, intuition universality is probably false.

This leaves open the question of *population-level* intuition universality. In every population (in some sense of “population” to be specified), do we find that a majority of people have a disposition or tendency to make problem judgments? In practice, discussions of universality of philosophical intuitions across cultures tend to focus on questions about populations, since individual-level universality in these domains is usually out of the questions. It is by no means obvious that problem intuitions are universal in this way. But it is also not obvious that they are not, and the matter is worth studying.

At the individual level, the most defensible universality thesis is perhaps *source universality*, which says that the hard problem has a universal source. According to this thesis, there is some factor that is present in everyone (or at least in every normal adult human), which tends to generate problem intuitions, and which explains problem intuitions when they are present. Individual-level source universality is consistent with the rejection of individual-level intuition universality, as long as there is some distance between the source and the intuitions: for example, perhaps the source needs to be taken up in a certain way to yield a disposition to make problem judgments.

In “The Meta-Problem of Consciousness” I said that it is an empirical question how widely shared problem intuitions are. I also said that I would work under the assumption that these intuitions are widely shared, or at least that they have a widely shared basis. That is, I assumed some version of intuition universality or at least source universality. I am still inclined to think that population-level intuition universality has a chance of being true for at least some problem intuitions, and that individual-level source universality is true. At the same time, it is clear that all this involves disputable empirical claims that are worthy of sustained investigation.

In the symposium on the meta-problem, a number of commentators question various sorts of universality. Some bring in empirical or textual evidence to make their case. Anna Wierzbicka uses considerations about human language. Justin Sytsma and Eyuphan Ozdemir use results from experimental philosophy. Zed Adams and Jacob Browning examine the history of philosophy.

Other commentators make related distinctions between two sorts of explanation of problem intuitions, expressing sympathy for approaches of a relatively non-universal sort. Graham Peebles contrasts non-inferential and inferential explanations. Adam Balmer contrasts hard-wired and soft-wired illusionism. Elizabeth Irvine contrasts deep and shallow explanations.

Some commentators offer specific explanations of problem intuitions that locate their source in non-universal factors. Hakwan Lau and Matthias Michel argue for a sociological explanation. David Rosenthal argues that problem intuitions arise largely from proto-theoretical prompting.

Some commentators focus on consequences of non-universality. Both Rosenthal and Asger Kirkeby-Hinrup both argue that if the problem intuitions are not universal, this tends to undermine the meta-problem challenge for theories of consciousness. Helen Yetter-Chappell argues that if there can be rational beings who lack problem intuitions, the standard anti-physicalist arguments will be undercut.

There is also some discussion of just how the hard problem should be understood, with Lisa Miracchi and Galen Strawson questioning the distinction between the hard problem and the easy problems of consciousness in different ways.

Despite all this, it is notable that the large majority of commentators who offer solutions to the metaproblem (discussed in “Solving the Meta-Problem of Consciousness”) seem to locate the source of problem intuitions in relatively universal factors. So there is at least an interesting disagreement here to address.

Which is correct? Do problem intuitions have a near-universal source, present in most normal humans and built in by evolution? Or do they have a shallow source, present in some people but not others, and deriving from culture? I cannot give a definitive answer here, but in what follows I will address a number of considerations that bear on these questions.

The history of the hard problem

Where does the hard problem of consciousness first appear in the philosophical literature? It is notoriously hard to find the hard problem, or even the mind–body problem more generally, in ancient and medieval Western philosophy. Wallace Matson initiated a debate about this in his 1966 article “Why Isn’t the Mind-Body Problem Ancient?”, where he answers the question by saying ancient philosophers didn’t have a concept directly corresponding to our concept of sensation, and where they did have concepts like this, they assimilated them to the body side of the mind–body divide. Peter King continued the discussion in his 2007 article “Why Isn’t the Mind-Body

Problem Medieval?” and gave a similar answer. Medieval philosophers such as Ockham had a clearer concept of sensation, but they still located it with the bodily rather than with the mental. Matson and King give at best an incomplete diagnosis of the title question, since the mind–body problem arises for thinking as well as for sensation and perception, but this is at least a start.

Of course one can find at least hints and echoes of the contemporary mind–problem in ancient and medieval philosophy. Here is Galen, from the second century CE:

A single body capable of sensation cannot be produced from many which are incapable of sensation, if the elements are incapable of being affected. ... But sensation certainly is of a different genus than shape, weight, or hardness, which belong to the atoms, or again than the others that belong to fire, air, earth, and water. ... Consequently, the body that is capable of sensation cannot be constituted either from atoms or from fire, air, earth, and water, so long as they remain unchanged and are just the sorts of things they are in accordance with the nature of each.

One also finds hints here and there in Indian and Tibetan philosophy (see Coseru 2018 for discussion). For example, here is Dharmakirti in the 8th century CE, in an exchange with his commentator Prajkaragupta, he considers the relation between consciousness and basic elements that underlie it:

If the effect, that is, consciousness, arises from these alone, that is, from the great elements without having its prior origin in life (that is, in an animating principle), why doesn't everything arise having the form of a living being? If the physicalist claims that this is so not because of the elements (taken individually, but rather their transformation), we are presented with the same question: Why doesn't this special transformation (of the elements as the cause for the arising of consciousness) occur everywhere?

Still, discussions of the mind–body relation (let alone the consciousness–physical relation) are surprisingly few. In the 17th century they start to become more common. The locus classicus for the mind–body problem is still Descartes's discussion in the *Meditations* (1641) and elsewhere. Descartes's conception of the problem differs in some ways from the contemporary problem of consciousness. First, the emphasis is not so much on explanation, which is central to the hard problem. Second, it is not clear that his conception of the mind corresponds to the contemporary

conception of consciousness. Among other differences, the focus is very much on thinking, with sense experience treated as secondary. Third, while he gave a central role to modal intuitions, the intuitions at play mostly concerned mind without body as opposed to zombie-like intuitions concerning body without mind.

There is a famous passage from Isaac Newton in a letter from 1672:

to determine more absolutely, what Light is, after what manner refracted, and by what modes or actions it produceth in our minds the Phantasms of Colours, is not so easie.

Sometimes just part of this passage (leaving out “what Light is, after what manner refracted”) is quoted, which taken alone would give Newton a reasonable claim to be the originator of the hard problem idea. In the fuller passage, the issues here are a little more diffuse, but certainly it is easy to understand the issue about phantasms of color as part of the hard problem.

In the *Monadology* (1714), Leibniz gives a classic statement of an explanatory-gap argument:

It must be confessed, moreover, that perception, and that which depends on it, are inexplicable by mechanical causes, that is, by figures and motions, And, supposing that there were a mechanism so constructed as to think, feel and have perception, we might enter it as into a mill. And this granted, we should only find on visiting it, pieces which push one against another, but never anything by which to explain a perception. This must be sought, therefore, in the simple substance, and not in the composite or in the machine.

The argument here is remarkably contemporary in form. There is some debate about whether Leibniz’s discussion concerns consciousness, as he allows that perception can be unconscious. On the other hand, it is arguable that Leibniz’s notion of consciousness was more like our notion of reflective or higher-order consciousness, and his notion of perception was more like our notion of phenomenal consciousness.

This gets at a more general issue complicating discussion of the problem of consciousness in the 17th and 18th centuries. In this period, “consciousness” was used mainly for something like reflective or introspective consciousness—that is, higher-order awareness of our own minds and mental states. There was (arguably) no real need for a term for phenomenal consciousness, since it was (arguably) taken for granted by most that every mental state is phenomenally conscious. If so, one should not look for references to “consciousness” for references to the problem of phenomenal

consciousness. The problem can be raised by ordinary terms for the mind and mental states. It is at least arguable that this is what Leibniz was doing here. Still, Leibniz did not elaborate and there was not a great deal of further discussion around that time.

For really definitive statements of the hard problem of consciousness, we have to wait until the 1860's with the classic statements of Thomas Huxley's "Djin" (1866), John Tyndall's "chasm" between physical processes and consciousness (1868), and Emil du Bois-Reymond's "ignorabimus" about the brain-consciousness relation (1872). Huxley was first, and cuts straight to the quick:

How it is that anything so remarkable as a state of consciousness comes about as a result of irritating nerve tissue, is just as unaccountable as the appearance of the Djin when Aladdin rubbed his lamp. (Huxley 1866)

The famous discussions by Huxley, Tyndall, and Dubois-Reymond show an entirely contemporary understanding of the hard problem of consciousness. At this stage the problem was off and running, and discussion continued with fits and starts until the present day.¹

Why did it take until the 1860's for the problem to be widely recognized in something like its current form? And what happened in the 1860's that changed?

There is an obvious answer. One has to take materialism seriously in order to take the hard problem seriously as a problem. If one is antecedently a dualist, the hard problem will be unsurprising and not especially worth addressing. The mental and the physical are fundamentally distinct, and that is that. One might like to know how they interact, but that leads us to other aspects of the mind-problem such as the interaction problem. The problem of explaining the mental in physical terms does not really arise.

Before the 19th century, few philosophers were materialists. Most philosophers were substance dualists, and most of the rest were property dualists. When philosophers such as Locke entertain the radical possibility of thinking matter, it was usually property dualism along with substance monism that was at issue. Perhaps there were occasional full materialists (Democritus and Hobbes, and perhaps Epicurus and La Mettrie?), but the view was rare. Without materialism as a foil, the hard problem did not really get a grip.

¹For more on the history, see Güzeldere 1997, Tennant 2008, and Leach and Tartaglia 2016. As Tennant notes, the problem was regarded as familiar even at the time: du Bois-Reymond apologizes for offering his audience "stale beer". It seems clear that despite the differences in focus and background, these theorists regarded their version of the mind-body problem as continuous with those of Descartes and Leibniz.

It was precisely in the 1860s that materialism started to become widespread. In Bertrand Russell's preface to Lange's book *The History of Materialism* (itself published in 1865), he describes this period as "the materialistic '60's". It was also around this time that the usage of "consciousness" for something like phenomenal consciousness became widespread. So it is no surprise that just at this time, the question of how consciousness can be explained in physical terms exploded.

This does not mean that before the 1860s, nobody had problem intuitions. On my analysis, many people had these intuitions. At the least, many or most philosophers and scientists would have been disposed to judge that mental states such as feeling pain are not physical and cannot be fully explained in physical terms. It is just that before the time of materialism, these judgments might be taken as so obvious to be hardly worth noting.

Of course I have not really provided a historical argument that problem intuitions are universal or even widespread. I have mostly made a case that some historical evidence that might seem to undermine the universality of problem intuitions is in fact entirely consonant with universality.

In their contribution to the symposium, Zed Adams and Jacob Browning use historical evidence to argue that a different sort of problem intuition is not universal. They are concerned with the inversion intuitions: the intuition that we can imagine that what looks red to us looks green to someone else, or that the sky might look to someone else the way that grass looks to us. In the target article, I said that inversion intuitions are "particularly robust". In *The Conscious Mind*, I suggested that few people lack these intuitions. Adams and Browning argue that I am wrong: historically, these intuitions emerge only in the 19th century, largely as a response to the discovery of color blindness.

Adams and Browning's strategy is mainly to separate two sorts of intuitions. They find plenty of historical evidence (in Sextus Empiricus, Gassendi, Rohault, Malebranche, Locke, and Poincare) of people entertaining *inversion intuitions*, where things look to have different colors to other people. They find little evidence of people entertaining *inverted twin* intuitions, which involve physical duplicates to whom things look to have different colors. Most of the people entertaining inversion intuitions in fact appealed to conditions such as jaundice where there are physical differences, so inverted twins are not really on the horizon. Inverted twins first make an appearance with Poincare in the early 20th century.

Adams and Browning say that my claims about the robustness and widespreadness of inversion intuitions concern inverted twin intuitions. This is not quite right. The relevant passages in *The Conscious Mind* and the target article both concern simple inversion intuitions. There is another point in *The Conscious Mind* where I invoke inverted twin intuitions to make an anti-materialist

argument, but the sociological claims are distinct and concern inversion intuitions. In the target article, only inversion intuitions are mentioned and inverted twin intuitions play no role. As far as I can tell, Adams and Browning's investigation supports the claim that inversion intuitions are at least fairly widespread.

Adams and Browning say that none of the relevant figures (before Poincare) use inversion intuitions to support problem intuitions; but it was no part of my claim that they did. Given our earlier discussion, it is no surprise that people did not argue this way before the 19th century, since physicalism was not at issue. The same point explains why inverted twins were not much of an issue. Now, Adams and Browning might say reasonably enough that inversion intuitions on their own are not enough to get the hard problem (qua problem for materialism) off the ground—one needs something closer to inverted twins for that. That is a fair point. At least, materialists have an easily available strategy for accommodating mere inversion intuitions. Still, inversion intuitions are a very good way of at least focusing attention on the phenomenally conscious properties of experience. If they are widespread, they may at least play some role in helping to focus people on the problem of consciousness.

The hypothesis that the discovery of color blindness helped to bring about inverted twin intuitions in Poincare and others is a fascinating one, but I am somewhat doubtful. First, colorblindness involves physical differences, so it is not clear why its status with respect to inverted twins is any different from jaundice. Second, theorists such as Poincare do not seem especially motivated by colorblindness. An equally plausible hypothesis is that the reason they consider inverted twins where others do not is tied to the rise of materialism.

Adams and Browning also give one counterexample to the universality of inversion intuitions: Aristotle, whose theory does not permit inversion. Aristotle is a strong realist about color, with redness and greenness as qualities out there in the world. Experiences of red resemble red things and rarely if ever go wrong. This makes it hard to have inversion: one would need cases where experiences do not resemble their causes. So certainly one could not have cases of *veridical* inversion on this view. One might think there could still be inversion as long as their can be color illusions, which are not obviously impossible or inconceivable. And even if Aristotle's theory forbids these things, a theoretical commitment to a claim does not entail that one cannot imagine it false, or wonder whether it is false. So I am not sure that we have enough evidence to rule out some underlying problem intuitions in Aristotle, masked by his theory.

If Adams and Browning are right that Aristotle and other naive realists do not have inversion intuitions, then intuition universalism will be false in this domain. Source universalism will remain

open: for example, someone could hold that the source of problem intuitions is a special sort of acquaintance with qualities that everyone has but that naive realists mistakenly take to be qualities of external objects.

Is the concept of phenomenal consciousness universal?

One aspect of the universality question is whether the concept of phenomenal consciousness is universal. That is, does every normal adult possess the concept of phenomenal consciousness?

In the target article I said that most subjects have concepts of specific phenomenal states such as feeling pain or experiencing colour, but that I am neutral on whether they also have a unifying concept of phenomenal consciousness. I am still inclined to be agnostic about this matter.

As I understand things, the concept of phenomenal consciousness need not be universal for problem intuitions to be universal. Universality of the concept at best makes a small difference for how the relevant problem intuitions are understood. If people have the generic concept of phenomenal consciousness (and of the physical), they they can presumably entertain questions such as whether consciousness is physical or whether it can be physically explained. If they only have specific concepts such as the concept of feeling pain, then the relevant problem intuitions will involve these concepts: for example, is the feeling of pain something physical, and can feeling pain be wholly explained in physical terms? Some phenomenal intuitions were already framed this way—for example, the knowledge intuition involving Mary concerns what it is like to see red—and it is not hard to reframe other intuitions in a similar way.

Anna Wierzbicka says that ordinary speakers of English have no such concept, and it remains a philosophers' construct. She says that English speakers have a concept of feeling (which covers both bodily feelings such as feeling toothache and thoughts such as feeling disappointment) and a concept of seeing (e.g. seeing a concert), but no unified concept that covers both feeling and seeing.

I am not so sure. At least in my dialect of English, the 'experiencing' covers both feeling and seeing. One can experience disappointment and experience a concert. I know that elsewhere Wierzbicka has questioned whether the English word 'experience' has counterparts in other languages, but for now I am just concerned with English. I think there is a prima facie case that this use of 'experience' at least in my dialect expresses a unified notion of phenomenal consciousness. I am not sure whether this is true in all dialects, though, and later in this section I will discuss evidence that may suggest that "experience" can have nonphenomenal readings too.

Another relevant word is 'aware'. I can reasonably be said to be aware of a toothache and aware of the dog, and perhaps aware of my disappointment, though it is perhaps not obvious whether these come to the same thing as feeling toothache and feeling disappointment. There are perhaps other uses of 'aware' to express knowledge that may or may not be conscious. But there is arguably at least a usage on which awareness is tied to phenomenal consciousness: if one is aware of something one is phenomenally conscious of it, and vice versa.

Wierzbicka also shows some sympathy with the "what it is like" formulation. Perhaps she will allow that there is something it is like to be disappointed, something it is like to see a dog, and so on. Perhaps not every ordinary use of this phrase covers phenomenal consciousness ("what was it like to grow up in Australia" doesn't obviously refer to phenomenal consciousness), but there is at least one fairly ordinary use that does.

I also wonder about the ordinary English word "mind". We certainly talk about thoughts and feelings of pain as being in one's mind. Perhaps it's not a common to speak of seeing as being in one's mind, but I think there is at least arguably a use that works this way. One issue is that "mind" may cover a lot which is not conscious, such as unconscious desires, background beliefs, and so on. So the fit may be imperfect, but again there is perhaps some concept that works this way.

In any case, I think that even if there is no ordinary English word that always and only expresses the concept of phenomenal consciousness, there are ordinary English expressions that have natural uses that express the concept of phenomenal consciousness. Of course it may be that my dialect has been corrupted and ordinary English doesn't work quite this way. I also accept that these words are at least not as common or as straightforward as Wierzbicka's favorites: see, hear, think, feel, want, and so on.

Moving beyond English, Wierzbicka notes that her Natural Semantic Metalanguage with 65 primitive words which are crosscultural universals includes SEE, HEAR, FEEL, THINK, KNOW, and WANT, but no generic term for phenomenal consciousness. I know she holds that English 'experience' does not have counterparts in all languages of the world, and perhaps she holds the same for 'aware' as well. On the other hand, Wierzbicka says explicitly that 'what it is like' is universally cross-translatable. If so we have at least a candidate for a universal expression for phenomenal consciousness, and a prima facie case for a universal concept of phenomenal consciousness (though as above someone might argue that the ordinary use of this phrase comes apart from phenomenal consciousness in some way).

Even in the absence of a single universal expression like this, there is also the option of translating talk of phenomenal consciousness into disjunctive talk of seeing, hearing, feeling, thinking,

and so on. Of course there will be questions. For a start, I wonder what happened to smelling and tasting and perhaps other conscious states that do not appear on the list. Knowing may well cover unconscious states and could be dropped from the list, and there are related questions about whether wanting, seeing, and so on can be unconscious (I think there is a case that on ordinary usage seeing and hearing are conscious states). Still, some version of this move might yield at least a disjunctive universal concept of phenomenal consciousness— though one might reasonably hope for more, as above.

Wierzbicka takes me to task for using obscure technical terms such as ‘phenomenal consciousness’ instead of plain English terms like ‘see’ and ‘feel’. I don’t apologize for using technical terms in an academic article. They play a key role in efficient communication in every discipline, including Wierzbicka’s discipline where technical terms like ‘prime’ and ‘universal’ are used all the time. That said it is nice to be able to translate one’s technical talk into nontechnical talk, and I have tried to do this on occasion.

Certainly, when we are trying to capture questions about the universality of certain intuitions and judgments, we need to be able to express those intuitions and judgments in nontechnical terms. I have offered a few ways to do that for general problem intuitions above. We also have the option of appealing to specific problem intuitions, which concern feeling pain or seeing red and the like. This gives us plenty of vocabulary for approaching questions about universality.

Justin Sytsma and Eyuphan Ozdemir present evidence from their studies in experimental philosophy to support the claim that the concept of phenomenal consciousness is not universal: ordinary people by and large lack the concept of phenomenal consciousness.

S&O say their non-universality thesis undercuts the assumption that problem intuitions are widespread and that they have a widely shared basis. As I am understanding problem intuitions, this inference is incorrect. Even if people lack problem intuitions using the concept of phenomenal consciousness, they may still have problem intuitions involving the concept of feeling pain, the concept of seeing red, and so on. And a basis for intuitions (where present) may be universal even if the intuitions and the concepts involved are not. Still, the non-universality thesis is interesting in its own right.

S&O’s studies, like Sytsma and Machery’s well-known study before them, focuses on attributions of mental states to human subjects and robots. Sytsma and Machery found that subjects were more willing to attribute ‘seeing red’ to robots and than they were willing to attribute ‘feeling pain’ and took this as evidence against a concept of phenomenal consciousness (since these two paradigmatic phenomenal states were treated differently). I suggested that the results can just as

easily be explained by a version of a hypothesis I put forward in the first chapter of *The Conscious Mind*. The idea is that while most mental terms have both phenomenal and nonphenomenal readings, 'feeling pain' strongly triggers a phenomenal reading while 'seeing red' does not. If so, and if subjects are inclined to attribute nonphenomenal but not phenomenal states to the robot, this explains the result.

S&O respond to my observation here by discussing the difference between 'pain' and 'feeling pain' (in the target article I said 'pain' where I should have said 'feeling pain'), but this does not come to much. More interesting is a set of follow-up studies where they use 'experience red' and 'experience pain' instead of 'see red' and 'feel pain'. This is a reasonable path as one might expect 'experience' to trigger a phenomenal reading more strongly than 'see'. I would have been antecedently inclined to expect that myself. Interestingly, while they do find this result, the overall pattern is similar to before. People are a little less willing to attribute 'experience red' to the robot than they are attribute 'see red', but they are a good deal more likely to attribute 'experience red' than to attribute 'experience pain' or 'feel pain'. They take this as strong evidence for the absence of a concept of phenomenal consciousness.

As before, I take this as evidence about usage of language. These results suggest that 'experience red' and 'see red' are less different than I thought, and that in particular that while 'experience red' may trigger a phenomenal reading a little more often than 'see red', it does so less often than 'feel pain' and 'experience pain'. This usage is quite compatible with subjects having a single concept of phenomenal consciousness.

S&O respond to this sort of suggestion by saying: if 'experience' does not express phenomenal consciousness, what does? I have offered some alternative suggestions above. My view is that the relation between words and concepts is far from one-to-one: most words in this vicinity have multiple meanings, and few words have a univocal reading. So even if no word univocally expresses phenomenal consciousness, this does not mean there is no concept of phenomenal consciousness.

S&O's reasoning also turns on the premise that if people have the concept of phenomenal consciousness, they will treat all phenomenal states alike. That seems a non sequitur. We have the concept of furniture and of physical object, but we do not treat all furniture alike or all physical objects alike. It is entirely consistent to hold that people have the concept of phenomenal consciousness and nevertheless are much more willing to attribute some phenomenal states (say, experiencing red) to robots than others (say, experiencing pain). I am far from sure that this is what is going on in this case and perhaps one can find evidence against it, but it is worth exploring.

That said, I remain open to the possibility that many people lack a general concept of phenom-

enal consciousness. People certainly seem to treat seeing, feeling, and thinking in very different ways. Even in the target article I noted that we find a more striking explanatory gap for seeing and feeling than for thinking, even when the thinking is phenomenally conscious. Following S&O, it could turn out that many people find a stronger explanatory gap for feeling pain than for seeing red. One reason might be that (as in the discussion of Aristotle) the qualities involved in seeing red seem more strongly to be outside the subject and so less problematic than those involved in feeling pain. Perhaps that is evidence that we do not have a unified concept, or perhaps it is just evidence that we treat different phenomenal concepts in different ways. This would not entail or even suggest that there is no hard problem of consciousness, but it would certainly enrich the psychology of the problem.

Are problem intuitions universal?

This leads to the closely connected question of whether problem intuitions are universal, at least at the population level. Here the methods of experimental philosophy (and related methods from experimental psychology and crosscultural anthropology) are highly relevant. In principle, these methods could be brought to bear on all our main classes of problem intuitions, though some classes are perhaps more promising for this purpose than others.

Explanatory intuitions (e.g. consciousness cannot be fully explained in physical terms) have been studied a little already (by Gottlieb and Lombrozo), but they raise tricky issues as there are many notions and standards of explanations. The notion especially relevant to the hard problem is a certain notion of reductive explanation, but it may not be easy to focus on it experimentally. Gottlieb and Lombrozo's work is in terms of whether science can explain various mental states—but this is a relatively weak claim, one that even a property dualist like me may be happy to accept. To get at the hard problem directly, one might ask whether neuroscience can fully explain various conscious states. But still different notions of explanation may be deployed in answering it. For example, even a dualist could say yes if they think causal explanation is at issue and neural processes cause consciousness. So it will take some work to focus the question correctly. And even then, the concepts are complex enough that I would not expect anything close to universality.

Metaphysical intuitions (e.g. consciousness is nonphysical) are more straightforward to state. For example, one can ask questions such as “Is feeling pain a process in the brain?”. The concepts involved are all straightforward though perhaps different people might still read the question as one about identity, dependence, or other forms of correlation. As I suggested in the target article

(and as Diaz Leon suggests), metaphysical intuitions may be weaker than some other problem intuitions, perhaps because they lie further from the source of problem intuitions than the other key intuitions here.

Modal intuitions (e.g. that we can imagine systems physically or functionally like us without consciousness) will also be tricky because of the many notions of imagination (conceivability, possibility) and the difficulty of articulating the right standard. As a result, I would expect experimental results to be far from universal. Still there would certainly be a great deal to be learned from the attempt.

Perhaps the most promising intuitions to study experimentally are knowledge intuitions. This is both because they are especially easy to state in plain language, and because they are among the more promising candidates to be universal intuitions. For example, we could ask whether Mary (who knows all about the physical processes in the brain but has never seen colors) knows what it is like to see red, or whether she learns what it is like to see red when she leaves the room. I would predict a very strong positive response to these questions. Of course the knowledge intuition in this simple form may not be enough to get one all the way to other forms of the hard problem, but it would at least be a start, and it might provide some guidance about where to look for universal sources of problem intuitions.

Of course simply asking subjects whether they agree with a statement will tend to underestimate problem intuitions, since a subject may have the intuition as a tendency but override it. An alternative might be to ask the subject whether they have an intuition or an inclination to accept the statement, even if they do not ultimately accept it. Perhaps this may still underestimate intuitions (since people may have a disposition without knowing it), but it may at least get closer. My prediction is that done this way, we would find a very high provenance of knowledge intuitions.

Of all the commentators, David Rosenthal expresses the most skepticism that problem intuitions are widespread. He notes that what I call problem reports are complex theoretical claims that are far from spontaneous descriptions of one's mental states and do not really deserve to be called "reports". Not much rests on the term, and I agree that problem reports are more complex than simple introspective reports that one is experiencing a mental state. He also suggests that they result from "background setup, prompting, and tacit proto-theoretic assumptions". That sounds plausible to me, but it is quite consistent with the claim that they are widespread. In any case, the widespreadness of problem intuitions is an empirical issue that I hope will be empirically tested before long.

Rosenthal suggests that "casting things in terms of a meta-problem [...] takes the problem

intuitions for granted, insulating them from challenge”. That seems wrong to me. Casting things in terms of the meta-problem if anything opens the door wide to illusionist views where the intuitions are challenged. Relatedly, Rosenthal says that I propose that the intuitions are accurate because this explains their being widespread. No such proposal appears in the article. I agree with Rosenthal that there is no straightforward inference from widespreadness to accuracy. Rosenthal also spends a good deal of time questioning whether problem intuitions are accurate and whether accepting them provides a useful constraint on a theory of consciousness. But it was really no part of my article (perhaps until the very end) to argue that problem intuitions are accurate or should be accepted. So this is a quite different debate, closer to the familiar debate about the hard problem than the meta-problem.

Perhaps Rosenthal means only that the meta-problem casting takes the *existence* of the problem intuitions for granted. That might be closer to the truth, but I note even if the intuitions are not widespread, there is still a meta-problem. The meta-problem then becomes the problem of explaining why the intuitions exist where they exist, and explaining the variability in these intuitions. Of course in this case it becomes easier to reject the intuitions, and thereby to hold that the meta-problem does not shed light on the hard problem. I take up some of these issues in the next section.

Consequences of non-universality

What are the consequences of accepting that universalism is false? One set of consequences is for the meta-problem challenge, which says that if a theory specifies a mechanism for consciousness, that mechanism must play a role in explaining problem intuitions.

Both Rosenthal and Asger Kirkeby-Hinrup note (in effect) that if source universalism is false (e.g. if problem intuitions are explained by special prompting and tacit proto-theoretical assumptions), then the meta-problem challenge is inapposite. The source of problem intuitions will not be consciousness itself but the special factors, and one would not expect a theory of consciousness to explain the problem intuitions.

This diagnosis seems right. I said in the target article that the meta-problem challenge was a challenge for realists rather than illusionists. I should have added that it is really a challenge for robust realists—those who think that the problem intuitions are mostly correct. If the intuitions are (noncoincidentally) correct, then they presumably have some connection to consciousness itself and a theory of consciousness should help to explain them. But if the intuitions are incorrect, then

it may be reasonable to hold that their source is independent of consciousness, so the meta-problem challenge need not be answered.

That said, my discussion of the meta-problem challenge was cast in terms of judgments and reports about consciousness generally, including simple phenomenal reports such as ‘I am conscious’ and not just problem reports. A central version of the meta-problem challenge applies to ordinary phenomenal judgments and reports, saying that what explains consciousness should also help explain these judgments and reports. Most realists about consciousness will presumably hold that these reports reflect at least something of the character of a subject’s conscious states. If so, it is reasonable to hold that a theory of consciousness can help explain phenomenal reports. For example, what brings about a phenomenal report may be some combination of the mechanisms of consciousness along with mechanisms of introspection and of speech. This version of the meta-problem challenge is apposite for any realist, not just for robust realists.

In the target article, I used Alvin Goldman’s argument against higher-order thought theory as an instance of this sort of meta-problem challenge. I said it is not clear how mere higher-order thoughts explain why we report mental states as being conscious (*prima facie*, that requires having higher-order thoughts about higher-order thoughts). Rosenthal responds with the interesting observation that outside special contexts (such as academic contexts) people do not spontaneously describe their mental states as conscious. Perhaps that is right. There are more than 200,000 Google hits each for “consciously think” and “consciously try”, but it is true that locutions such as “consciously see” are less common (37,000 hits). Still, Goldman’s point really concerned judgments and not reports: when a mental state is conscious, we are usually in a position to know that it is conscious. One aspect of the meta-problem challenge for higher-order thought theories is to provide a good explanation of that.

Kirkeby-Hinrup objects to the meta-problem challenge, suggesting that the mechanism of consciousness may be content-neutral. That is, it takes any content and makes it conscious, as a higher-order thought does, for example. In this case Kirkeby-Hinrup suggests that we should not expect judgments about consciousness (which are just contents) to shed any light on the mechanism. This seems wrong to me. It is true that if a mechanism makes a content like “Red square there” conscious, we should not expect that content itself to shed any light on the mechanism. But as a result of the content becoming conscious, we would expect new judgments with new contents such as “I am conscious of a red square there” as well as potential problem judgments and the like. Even if the mechanism is content-neutral, we expect that these new judgments will be caused by the mechanism’s being applied to the original content. So for a theory of consciousness to be plau-

sible, it needs to be able to tell a story about how the mechanism and the original content together lead to these judgments. For example, the higher-order thought theory needs to be able to tell a story about how higher-order thoughts plus a “Red square there” content bring about judgments such as “I am conscious of a red square there”. That is a version of the meta-problem challenge.

Finally, Helen Yetter-Chappell argues that if certain “unperturbed agents” who rationally lack problem intuitions are possible, this undercuts the dialectical force of arguments against physicalism. The idea is that those beings will rationally not find zombies conceivable, so there is a case for zombies being inconceivable after all. The anti-physicalist will respond by saying that these unperturbed agents are not ideally rational, as an ideally rational agent will find zombies conceivable. Yetter-Chappell suggests that this involves ad hoc special pleading that leaves anti-physicalism in a dialectical balance. I’m not so sure about this. Almost any a priori claim could be doubted by some rational being (think of the way the principle of non-contradiction is doubted by dialetheists or the infinitude of primes is doubted by finitists). I don’t think the mere possibility or even the actuality of such beings does much to undermine our own a priori justification for the claims in question— though perhaps their reasons will provide a case that we will need to answer.

Yetter-Chappell goes further by sketching a possible design for an unperturbed agent. This agent has automatic subpersonal matching between phenomenal concepts and physical concepts, so it transitions directly between claims about pain and claims about correlated neural activity (TNA). The agent thereby judges that pain is TNA, rejects the conceivability of zombies, and so on. Now, it is not obvious to me why a mere subpersonal matching process should lead to these judgments. Prima facie it looks like a strong empirical association that can conceivably fail, like most empirical associations. And even if these transitions are psychologically irresistible, I don’t think they will be justified a priori. So if the agent uses their psychological force to reject conceivability (a priori consistency) intuitions and the like, I think they are irrational or at least not ideally rational. Of course it is possible to design agents who are irrational in this way, but as with philosophers who reject a priori claims, their existence doesn’t do a great deal to undermine our own justification for these claims.

Introspection and inference

Graham Peebles asks whether problem judgments are deliverances of introspection alone, or whether they require inference? That’s easy. They are not deliverances of introspection alone. As Peebles notes, problem judgments involve concepts such as explanation, knowledge, and phys-

icality, which are not plausible among the contents of introspection of conscious states in general. Introspection of conscious states largely delivers us with truths such as “I am having such-and-such conscious state” or perhaps “This experience has this property”. None of this yet is the territory of problem judgments which are far more complex. To get to the problem judgments, more work is required.

Some problem judgments may be a priori, and therefore not strictly speaking grounded in introspection. For example, “consciousness cannot be explained physically” does not obviously need any introspective justification. The same goes for many others which simply have a generic reference to consciousness or to feeling pain or something similar. Here, the role of introspection is to give one the concept of consciousness or of feeling pain in the first place. Once one has these concepts and the concept of the physical, the judgment is at least a natural one to make on a priori grounds.

Other problem judgments involving demonstrative concepts may be empirical and grounded in introspection. For example, when I judge “Mary could not know what *this* is like”, it is plausible that introspection plays a role. Introspection tells you that you’re having the experience, and further reflection allows you draw further conclusions. When this happens, I think that as Peebles suggests there is inference from the contents of perception, perhaps along with other largely a priori premises and a priori reasoning.

Some problem judgments may be quite close to the contents of introspection. On some views it can be part of the content of introspective judgment that “I am acquainted with R”, or some such. Then the acquaintance intuition “There are qualities I am acquainted with” is only a tiny step beyond the content of introspection. Likewise maybe “there are qualities I know about” is close to this content, and other knowledge intuitions are not far behind. But still I would say that a minor inference is required.

This role for reasoning does not entail that intuition universalism is false. Some sorts of reasoning seem universal. For example, it is arguable that almost everyone has at least a disposition to judge that there is something in front of them when they have a visual experience as of something in front of them. It is not out of the question that the small steps of reasoning requires to get from introspection to intuitive judgments are like this. And obviously, this role for reasoning is perfectly compatible with source universalism. On a natural version of source universalism, introspection provides the source, which is present in everyone. Reasoning provides the step from the source to intuitions, which may or may not be present in everyone.

Source universalism, realism and illusionism

Source universalism is especially natural for a realist about consciousness such as myself who accepts many problem intuitions. As a realist, I suspect that the source of problem intuitions lies in consciousness itself along with certain perceptual and introspective mechanisms that are likely to be universal. Source universalism is also common among illusionists, however, as numerous illusionist treatments in the symposium bring out.

To avoid trivializing source universalism, more needs to be said about just what the conditions are for some factor to be a “source”. Any number of universal factors (e.g. perception and introspection) will play some causal role in problem judgments, and if these count as “sources” then source universalism will hold trivially. SO sourcehood requires more. To a first approximation, we can think of the source of problem intuitions as the primary explanatory factor in explaining the intuitions when they are present, possibly combined with a secondary factor which accounts for the step from source to intuitions. Of course different theorists may hold different views of what counts as the primary explanatory factor. It is these differences that explain why some theorists are source universalists and others are not.

The issue of source universalism vs. variabilism interacts in interesting ways with the issue of realism vs. illusionism. We have seen that it is natural for a robust realist to be a source universalist. Perhaps there may be exceptions: epiphenomenalists who deny that consciousness is the source of problem intuitions, or realists who hold that culture drives problem intuitions all the same. Some non-robust realists (like Rosenthal) may hold that while consciousness exists, the problem intuitions are largely mistakes generated by bad reasoning. But speaking for myself, robust realism about consciousness is one major reason why I favor source universalism.

For an illusionist about consciousness, the issue is not quite so clear. Most of the illusionists in the symposium appear to be source universalists. Illusionists often hold that the illusion is strong and deep-seated, and deep-seated intuitions are more naturally regarded as universal. But an illusionist can also hold that the illusion is relatively shallow and arises from variable factors. We might talk about universalist illusionism and variabilist illusionism here, just as one might talk about universalist realism and variabilist realism.

Adam Balmer makes a useful distinction between hard-wired illusionism, where the illusion is the result of natural selection, and soft-wired illusionism, where it is the result of culture. Balmer’s distinction is useful beyond the realm of illusionism alone. We can also talk of hard-wired intuitions (produced by evolution) and soft-wired intuitions (produced by culture). In a similar fashion

we can also talk about hard-wired and soft-wired realism.

Balmer's hard-wired/soft-wired distinction is not quite the same as the universal/variable distinction, but it is strongly correlated. By and large, one would expect universalists to hold that problem intuitions (or at least their sources) are produced by evolution, and variabilists to hold that they are produced by culture. The correlation need not be perfect, though, as many factors produced by evolution (hair color, say) are variable, and some factors produced by culture (tool use, say) are near-universal.

Balmer observes that my framing of the issues call for a sort of alliance between realists and illusionists in endorsing a hard-wired view. He argues that this makes sense for realists, in effect because soft-wired realism is not a tenable option. But he thinks illusionists should reject the alliance and accept the soft-wired view, as hard-wired illusionism is problematic.

On soft-wired realism: Balmer suggests that it is absurd for a realist to say that we only noticed consciousness at the time of Descartes, if consciousness was there all along. I agree that this is at least somewhat absurd where phenomenal intuitions are concerned ("I am conscious"), but perhaps it is less absurd where problem intuitions ("Consciousness is hard to explain") are concerned. Still, for a realist it remains plausible that at least the source of problem intuitions is hard-wired, even if not the intuitions themselves. If we understand the soft-wired view as a view about the source of problem intuitions, I agree that soft-wired realism is not an especially attractive option.

On hard-wired illusionism: one reason Balmer gives to reject this view is that illusionism requires us to believe that phenomenal judgments involve a serious mistake, and it is easier to see how such a mistake might be soft-wired (produced in the history of philosophy) than hard-wired (produced by evolutionary history). I am not so sure. It is not too hard to see various cases where mistakes may be evolutionarily advantageous. Relatedly, he says that given a belief in something that doesn't exist, our first port of call should always be the history of ideas. We should look for hard-wired mechanisms only if the idea is "especially resistant to refutation, or pervasive across culturally isolated populations". Again I am not sure about the general principle. But in any case, *if* problem intuitions are hard-wired then their source is pervasive across populations. It is not clear that Balmer has given us any reason to think that they are not pervasive, even if illusionism is true. And if they are pervasive, then by Balmer's lights we can reasonably look for hard-wired mechanisms.

Balmer also says that hard-wired illusionism rests on an "implausible just-so story". He illustrates this point mainly with a critique of Nicholas Humphrey's hard-wired illusionism, saying that Humphrey's suggest that the illusion makes things "magically delightful" is an unreliable and

inefficient way of getting us to value and preserve those things. I am not sure that Balmer's point generalizes beyond Humphrey's account, however. In the case of color, it is easy to see how illusory representation of objects as having primitive colors might be a highly efficient way to get us to be sensitive to the relevant distinctions between objects. Not keeping track of the underlying physical properties may well enhance efficiency. The same may apply to introspection's representing us as having primitive mental qualities. It might be a highly efficient way for the mind to keep track of the mind. So it is not clear why hard-wired illusionism must lead to inefficiency.

Balmer says that illusionists should reject the project of finding a topic-neutral explanation of problem intuitions. I think they need not reject this. Suppose we have an explanation of problem intuitions according to which consciousness is an illusion. Then one would expect that there will be some topic-neutral component of that explanation (an algorithmic story, say) that is neutral on the existence of consciousness. The same point applies to realists, who think the correct (and perhaps best) explanation involves consciousness but can still endorse explanations put in neutral terms. In this way the topic-neutral project remains a project for everyone.

Elizabeth Irvine makes a related distinction between "deep-seated" and "shallow" explanations of problem intuitions, where the former are in terms of fairly basic cognitive mechanisms and the latter in terms of something at a higher level. As with the hard-wired/soft-wired distinction, this distinction seems likely to correlate with the universal/variable distinction even if the correlation is not perfect.

Irvine focuses especially on what explains the variability across subjects in problem intuitions. She suggests that this variability is harder to explain on a "deep-seated" view of problem intuitions than on a shallow view, and she suggests that this might give some support for illusionism.

Irvine starts with the hypothesis that some people may not have problem intuitions at all. She does not endorse the hypothesis, but notes that some empirical evidence at least suggests it. She says this hypothesis is hard to reconcile with a deep-seated view of problem intuitions. If problem intuitions arise from deep-seated cognitive machinery, why should they differ from people?

Now, it is clearly true that if some people lack problem intuitions, then intuition universalism is false. We can then rule out that intuitions arise from universal cognitive machinery in the sense that the machinery guarantees the presence of the intuitions. At the same time, the absence of problem intuitions in some people is clearly compatible with source universalism. The source of our intuitions could be tied to universal cognitive machinery. It is just that if the intuitions themselves are variable, the pathway from the source to the intuitions must involve non-universal factors.

For example, a natural story here is that problem intuitions arise from universal introspective processes plus variable reasoning processes. Perhaps introspection universally represents ourselves as having certain primitive properties, or as being acquainted with certain primitive properties. From here, it is natural to reason on the basis of introspection that we have special knowledge of these properties or that these properties are irreducible. Even though these judgments are based in introspection and may well be justified by introspection, they go beyond introspection and need not be universal. Likewise, even a tendency to make these judgments when prompted may not be universal. This part of the process relies on reasoning mechanisms that may be widespread but not universal.

A picture like this can help resolve the “tension” Irvine sees for a realist who holds that problem intuitions are based in consciousness (which is universal) but are not universal themselves. A strong realist position demands a close connection between consciousness and our judgments about it, as Irvine says, but the connection need not be so tight as to be universal. For similar reasons, there need not be a tension for a hard-wired illusionist who thinks the source of the illusion is universal but the intuitions themselves are not. Source universalism without intuition universalism is a viable combination for realists and illusionists alike.

Of course all this leaves open just what explains the variability in problem intuitions, if indeed they vary. I don’t have a strong view about this and agree with Irvine that it is an important question. If something like the picture above is correct, one would expect these to be factors that affect reasoning and inference rather than the basic mechanisms of introspection—though there may well be less basic elements in introspection that are variable too.

If we restrict our attention to variations in problem judgments and problem intuitions among theorists, many factors may be at play, including the theorists’ commitments, their training, and their philosophical personality more generally. As for variations in ordinary people, numerous hypotheses are available. One hypothesis I have sometimes entertained is that some people are more “transparent” or outward-directed thinkers, and have trouble attending to the qualities of their conscious experience, while others are more inward-directed and easily attend to their mental lives. It would not be entirely surprising if those in the latter class have problem intuitions more often than those in the former class. The matter is ripe for empirical investigation.

Sociological explanations

The clearest case of a soft-wired explanation of problem intuitions is provided by Lau and Michel (Lau and Michel), who offer a “socio-historical” explanation of why the hard problem of consciousness has persisted.

Their explanation has two parts. First, the reason the hard problem has persisted is that we have not yet developed empirically adequate theories of consciousness. Second, the reason we have not yet developed empirically adequate theories of consciousness is because of a “guru effect” in the field where the dominance of gurus endorsing revolutionary approaches leads to a breakdown in the normal scientific process.

Lau and Michel spend the greatest part of their article by far defending the second premise, which does not have much to do with the hard problem. I am not sure whether I find the premise plausible. They note that the guru effect is far weaker in Europe than in the US, but empirical theories of consciousness are still not in an advanced state there. An alternative hypothesis is that we have not yet developed empirically adequate theories of consciousness because of in-built limitations in the area compared to other areas of neuroscience, including especially difficulties of measurement. But assessing this matter is obviously a nontrivial project.

In any case, it is the first premise which is most relevant to the meta-problem. Lau and Michel assert it but do not try to defend it. For what it is worth, I find this premise implausible. I am inclined to think that when we have empirically adequate theories of consciousness, this makes problem intuitions stronger rather than weaker.

To see this, note that by an empirically adequate theory of consciousness, Lau and Michel do not mean a complete theory of consciousness that solves the hard problem. They mean something more like a good theory of the neural correlates of consciousness, which tells us the conditions under which consciousness is present in humans and other animals.

In my experience, theories of this type tend to accentuate the hard problem rather than remove it. Without an empirically adequate theory, we can hope that unspecified future discoveries in neuroscience might close the gap. But once we have an empirically adequate theory with neural correlates of consciousness, it is typically all the clearer that there is a gap between these processes and consciousness. In this case we will have done most of the relevant neuroscience so the appeal to the future does not carry much weight. Indeed, philosophers and lay people who think about the hard problem often assume for the sake of argument that we have an empirically adequate theory of consciousness. This does little to remove the problem.

Maybe Lau and Michel think that a truly empirically adequate theory would have some special feature that helps cross the explanatory gap and remove the hard problem. But that now gets us back to the revolutionary territory that Lau and Michel wanted to avoid. Or perhaps having reliable correlations and predictions in hand would tend to remove the sense that there is a problem. This might work for people who were largely moved by the problem of correlation and prediction in the first place, but for those who are moved by the hard problem, it is hard to see how correlation and prediction helps.

Late in their article, Lau and Michel suggest that a focus on the hard problem itself may have slowed down progress in the science and thereby prevented the dissolution of the hard problem. This version does not ring true either. For a start, very few people working on the science focus on the hard problem. Perhaps the idea is that the mere acknowledgment of the hard problem can prevent people from going into the field and thereby slow down progress. I don't think scientists are this resistant to a challenge, though. Far more common are people who got into the field because of the hard problem and ended up doing empirical science. In any case, it remains as implausible as ever that if we had only focused more on the easy problems, we would then have solved the hard problem.

What is the hard problem?

Finally, a couple of commentaries address the nature of the hard problem itself, so it is worth saying a few words about the problem as I understand it. It might be put in two words, as Tom Stoppard did in his play *The Hard Problem*: explain consciousness! In the target article and various other works I have put it more mundanely: why and how do physical processes in the brain give rise to conscious experience? One could also stay neutral on the priority of physical processes (which idealists may question) by asking simply: why and how does conscious experience arise? These questions include both the problem of explaining why consciousness exists at all, and the problem of explaining the specific character of specific experiences.

The hard problem of consciousness contrasts with the easy problems of consciousness: explaining reportability, perceptual discrimination, integration, the difference between wakefulness and sleep, and so on. These are not problems of phenomenal consciousness, but they are problems of consciousness in other uses of the term (access consciousness, for example). Galen Strawson rejects this usage because he thinks only phenomenal consciousness deserves to be called "consciousness", but this is a verbal issue. Certainly I am happy to say that the hard problem of con-

consciousness is *the* problem of explaining phenomenal consciousness. And certainly it is sometimes useful to contrast the hard problem with easy problems that are not problems of consciousness in any sense, such as problems of explaining ordinary behavior.

Lisa Miracchi rejects the distinction between the hard problem and the easy problems, saying the hard problem is not so hard and the easy problems are not so easy. She argues that we should understand the problem of consciousness as the *generative question* of what gives rise to consciousness and what the principles of generation are. Miracchi notes that the generative question does not require answering the “nature question” of what consciousness is, it does not require giving a transparent explanation of consciousness that shows why consciousness has to exist (given underlying physical processes), and it does not require solving the mind–body problem. The question can be answered by ordinary tools of cognitive science that find and systematize the principles by which consciousness depends on the brain.

I agree that the generative question is an important one, but in making it easier than the mind–body problem Miracchi has also made it easier than the hard problem as I understand it. Miracchi herself notes that generative explanations are weaker than the transparent explanations and nature explanations that we find elsewhere in science and that many require for the hard problem. So in moving from the hard problem to the generative problem, Miracchi is not really arguing that the hard problem is not so hard. She is instead noting that there are important problems that are easier than the hard problem, and recommending that we focus on those.

It is true that the best scientific theories of consciousness that we have are generative theories. But in many other areas of science, we can do better. In *The Conscious Mind* and elsewhere, I argued that for most phenomena in science, transparent reductive explanation is the gold standard, and transparent reductive explanations in physical terms are available in principle. It is not unreasonable to hope for such an explanation for consciousness.

Now, my own view is that a transparent explanation of consciousness in physical terms is impossible. If so (and if other transparent explanations are also ruled out), some sort of generative explanation appealing to fundamental psychophysical laws may be the best we can do in solving the hard problem. But that will be a lowering of our sights. To see this, note that generative explanations do not really explain why consciousness exists in the first place. They assume that consciousness exists, and explain its character in specific cases by appeal to associated physical processes. I have also argued that if the principles of generation themselves cannot be transparently explained, they must be taken as fundamental, which leads to a radically nonreductive theory of consciousness. Miracchi may disagree, and I am sure there is more to say. In any case, I agree that

generative explanations of consciousness are both important and viable. But that does not make the hard problem easy.

A residual question: on my understanding, is the hard problem then the problem of *transparently* explaining consciousness, of *generatively* explaining consciousness, or something else? It is not quite the former, as I think there can be solutions to the hard problem in non-transparent terms: for example, theories that posit fundamental psychophysical laws. It is not quite the latter, as the latter is a relatively easy problem and we can hope for more. Perhaps it would more accurate to say that the hard problem as I understand it is the problem of transparently explaining how consciousness arises, if that is possible, and of non-transparently explaining how it arises, if transparent explanation is impossible. I think something like this is captured by the standard formulation which includes both “how” (suggesting a generative element) and “why” (suggesting a request for transparency if possible). In any case, the most important issue is not the formulation, but the substantive issue of just what sorts of explanation are and are not possible.

Miracchi also argues that the easy problems are not so easy, since all the “easy problems” (including the meta-problem itself) involve consciousness at least in paradigm cases. Here Miracchi echoes a line taken by E.J. Lowe in his 1995 commentary “There are no easy problems of consciousness” and by David Hodgson in his 1996 commentary “The easy problems ain’t so easy”. I would reply much as I did in my 1997 response, by conceding the point that these phenomena often involve consciousness and arguing that there are functionalizable explananda in the vicinity to which the main points apply.

The same point applies to the meta-problem itself. In the target article I recognized that problem judgments may involve consciousness, and argued that there is a topic-neutral explanandum in the vicinity that can be topic-neutrally explained. For example, we have propensities to make certain noises and inscriptions, naturally interpretable as concerning consciousness. That is a property that we in principle share with zombies and that does not require consciousness. Maybe Miracchi will hold that this is no longer the original meta-problem of explaining our judgments, and there is something to this. Nonetheless, it is in principle a tractable problem in cognitive science and one that is open to explanation using familiar tools. If we want to focus on the original problem, we can always explain it *modulo consciousness*: that is, explain problem judgments up to but not including their conscious components. This may not be a full explanation of the judgments, but it will still be an enlightening project to engage in.

References

- Adams, Z. & Browning, D. 2020. How color qualia became a problem. *Journal of Consciousness Studies*.
- Balmer, A. 2020. Soft-wired illusionism versus the meta-problem of consciousness. *Journal of Consciousness Studies*.
- Chalmers, D.J. 1997. Moving Forward on the Problem of Consciousness. *Journal of Consciousness Studies* 4:3-46.
- Coseru, S. 2018. Consciousness and the Mind-Body Problem in Indian Philosophy. In (R. Gennaro, ed.) *Routledge Handbook of Consciousness*. New York: Routledge.
- Du Bois-Reymond, E. 1872. ber die Grenzen des Naturerkennens. Leipzig: Veit and Co. Translated as "Limits of our knowledge of nature". *Popular Science Monthly* 5:17-32, 1874.
- Güzeldere, G. 1997. The many faces of consciousness: A field guide. In (G. Güzeldere, O. Flanagan, and N. Block, eds.) *The Nature of Consciousness*. MIT Press.
- Hodgson, D. 1996. The Easy Problems Ain't so Easy. *Journal of Consciousness Studies* 3:69-75.
- Huxley, T. H. 1866. *Lessons in Elementary Physiology*, London: Macmillan & Co..
- Irvine, E. 2019. Explaining Variation within the Meta-Problem. *Journal of Consciousness Studies* 26 (9-10):115-123.
- King, P. 2005. Why Isn't the Mind-Body Problem Medieval? In *Forming the Mind*. Springer Verlag.
- Kirkeby-Hinrup, A. 2020. Two caveats to the meta-problem challenge. *Journal of Consciousness Studies*.
- Lange, A. 1879. *The History of Materialism*. New York: Routledge.
- Lau, H. & Michel, M. 2019. A Socio-Historical Take on the Meta-Problem of Consciousness. *Journal of Consciousness Studies* 26 (9-10):136-147.
- Leach, S. & Tartaglia, J. 2016. *Consciousness and the Great Philosophers: What Would They Have Said About Our Mind-Body Problem?* Routledge.
- Lowe, E.J. 1995. There Are No Easy Problems of Consciousness. *Journal of Consciousness Studies* 2:266-71.
- Matson, W.I. 1966. Why isn't the mind-body problem ancient? In (P. Feyerabend & G. Maxwell, eds.) *Mind, Matter, and Method: Essays in Philosophy and Science in Honor of Herbert Feigl*. University of Minnesota Press.
- Miracchi, L. 2019. None of These Problems Are That 'Hard'... or 'Easy': Making Progress on the Problems of Consciousness. *Journal of Consciousness Studies* 26 (9-10):160-172.

Peebles, G. 2020. Two problems for non-inferentialist view of the meta-problem. *Journal of Consciousness Studies*.

Rosenthal, D. 2019. Chalmers' Meta-Problem. *Journal of Consciousness Studies* 26 (9-10):194-204.

Strawson, G. 2019. Underestimating the Physical. *Journal of Consciousness Studies* 26 (9-10):228-240.

Sytsma, J. & Ozdemir, E. 2019. No Problem: Evidence that the Concept of Phenomenal Consciousness is Not Widespread. *Journal of Consciousness Studies* 26 (9-10):241-256.

Tennant, N. 2008. Mind, mathematics and the Ignorabimusstreit. *British Journal for the History of Philosophy* 15:745-773.

Tyndall, J. 1868. Scientific Materialism. President's Address to the Mathematics and Physical Section of the British Association.

Wierzbicka, A. 2019. From 'Consciousness' to 'I Think, I Feel, I Know': A Commentary on David Chalmers. *Journal of Consciousness Studies* 26 (9-10):257-269.

Yetter-Chappell, H. 2019. Idealization and Problem Intuitions: Why No Possible Agent is Indisputably Ideal. *Journal of Consciousness Studies* 26 (9-10):270-279.