On the Meta-Problem

Joseph Levine, University of Massachusetts Amherst

According to Chalmers, the meta-problem of consciousness is "the problem of explaining why we think that there is a problem of consciousness" (pg. 6). Chalmers marches through quite a few distinctions and positions in the search for a "topic-neutral" solution to the meta-problem. Though he doesn't argue for a particular solution, or position, on either the first-order problem or the meta-problem, in this paper, he does say at one point:

"My own tentative view is that the most promising solution to the meta-problem lies in primitive relation attribution and the sense of acquaintance: our experiences seem to primitively acquaint us with qualities in the environment, and these experiences are themselves objects of acquaintance. I favor a realist theory of consciousness where consciousness does in fact involve acquaintance in this way." (pg. 39)

If I understand him correctly, I basically agree with what he says in this passage. I do think the key to understanding both consciousness itself and addressing the meta-problem is a matter of understanding what acquaintance is and what its objects are. Unfortunately, I think there are still some serious mysteries lurking here, which I will try to present briefly in this commentary.

But first some initial set-up. For one thing, it seems to me that the meta-problem looks very different depending on one's position regarding the relation between consciousness and physical-functional properties. If one is a materialist who acknowledges that there is, or at least seems to be, an explanatory gap between the physical-functional and the phenomenal, the meta-problem is straightforward. When we contemplate identifying phenomenal states with physical-functional states, we hit a cognitive block; somehow, it just doesn't "compute". We find ourselves wondering how this - pointing to our current phenomenal experience - could possibly just be the same thing as electro-chemical activity in the brain, or the workings of a computational process. Put the other way around, there doesn't seem to be an explanation of why there is something it is like to occupy these physical-functional states - or an explanation of just what it is like to occupy them - as opposed to these other ones. If everything really is physically explicable, then why don't we have an idea (of?) what an explanation of what it is like would look like?

On the other hand, if one isn't a materialist, and one doesn't think there really is an explanation of what it is like to occupy phenomenal states in physical-functional terms, then the response to the meta-problem should be obvious: we find a problem because there is one! We have these intuitions of an explanatory gap because there is a gap in nature, because one cannot in fact explain what it is like in physical-functional terms. Unfortunately, as Chalmers makes clear in the paper, the anti-materialist can't get off so easily. The meta-problem is also a problem for them; not because, as with the materialist, they have to explain why what isn't so appears to be so, but rather because, even if it is so, it's not obvious how to explain its seeming

so to us. When you endeavor to marry anti-materialism with plausible models of cognition, it becomes unclear how the special, non-physical nature of conscious experience could rightly be attributed a role in our coming to believe, or judge, or intuit that conscious experience isn't physical.

So the meta-problem plays out quite differently depending on one's position on the first-order problem, the relation between the physical-functional and conscious experience. Let me briefly comment on the materialist side of that coin first, though my main objections to the relevant moves have already been made elsewhere, particularly in Levine (2018, chapter 1). As I put the problem in that paper, there seems to be a "core contrast" between our cognitive responses to standard theoretical identities, like "water is H_2O", and identities between phenomenally conscious states and physical-functional states. Though the former are empirically determined and clearly defeasible, they make sense to us and provide a wide range of explanations of the phenomena involving those natural entities and kinds. These standard theoretical identities are intelligible to us; we see clearly how they can be true, even if some may doubt their actual truth.

But when it comes to identities like "pain is the firing of c-fibers", or "the experience of seeing red is having one's R-G channel in the R range", we find the alleged identities baffling. Why is it like <u>this</u> – pointing to the reddish feature of one's visual experience – to have these neurons firing this way as opposed to those firing that way? Why is it so hard to see phenomenal properties as a species of physical-functional properties?

Materialists have proposed numerous solutions, many of which Chalmers discusses. There are basically two lines to take here: the first is to argue that phenomenal experience is real but somehow conceptualized in a way that causes a kind of cognitive disconnect between the ways we conceive features of our own experience and physical-functional features. To put this another way, representations of phenomena conceived through the first-person perspective are in some way prevented from integrating with representations of phenomena conceived through the third-person perspective. This lack of integration is then cognitively interpreted as unintelligibility or puzzlement.

The other line to take is that phenomenal experience as we conceive it from the first-person perspective is a deep, probably unavoidable illusion. What we think are the qualities of experience are not in fact qualities of anything. Phenomenal experience being an illusion would solve the meta-problem because it would tell us that we can't seem to accept mind-body identities (or realization relations) because no real phenomenon does, or even could (according to some) answer to our first-person conception of experience. Rey (1995) compares our conceptions of phenomenal experience with conceptions of magic or incompatibilist free will. No characterization of these latter two phenomena in physical-functional terms would be met with acceptance by those holding these conceptions, but that's because nothing in the real world exists that answers to them. The claim is that the same goes for phenomenal experience.

As Chalmers says about himself in this article, if I were to be convinced of materialism, "illusionism" would be my favored response to the meta-problem; though I do think it suffers from serious defects. Here's why. To go the first route, essentially the "phenomenal concepts strategy", requires that one find a way of characterizing phenomenal concepts (or, what I described above as "representations of phenomena conceived through the first-person") so as to both attribute to them genuine satisfaction conditions that can be captured in physical-functional terms, and also account for their inability to be epistemically integrated with standard third-person concepts of these physical-functional properties. The crucial point, however, is that whatever features of phenomenal concepts it is that one appeals to in explaining this peculiar epistemic character must themselves not be subject to an explanatory gap. What makes phenomenal concepts special has to be itself a matter of physical-functional properties, or else materialism isn't vindicated by the phenomenal concepts strategy. As I argued in Levine (2018, chapter 1), it doesn't seem to me that materialist advocates of phenomenal concepts have successfully articulated a position that meets what I called "the materialist constraint".

Not wanting to repeat that argument in detail here, let me just say briefly how it goes. In order to really explain what I called above "the core contrast" – why mind-body identities provoke cognitive dissonance in a way standard theoretical identities don't – it seems to me you need to posit some special relationship holding between phenomenal concepts (or the mental states involving them) and the phenomenal properties, or experiences, they are about. But the only kind of relation that seems up to the task is something like a traditional acquaintance relation, which is itself as mysterious from a materialist standpoint as are the phenomena with which we are allegedly acquainted. Thus even if one thinks one has addressed the meta-problem – why there is an explanatory gap – by appeal to phenomenal concepts, another explanatory gap opens up when considering the special nature of these concepts.

Prima facie one doesn't have the same problem with illusionism. After all, one needn't posit any special relation between the relevant phenomenal concepts and the phenomena they purportedly pick out since, on this view, there are no such phenomena picked out by these concepts. But illusionism doesn't get off the hook that easily. For one thing, as mentioned by Chalmers, illusionism suffers from a rather serious case of unbelievability. "Really", one wants to say, "there is no conscious experience?" True, illusionists have their account of why this seems so unbelievable, but it's hard for any such account to surmount the initial incredibility of the doctrine. Second, as Chalmers cites Kammerer's (2018) "illusion meta-problem" in this regard, it's not clear that the illusionist can really account for the fact that we have this persistent illusion in terms that satisfy the materialist constraint (not his way of putting it).

One way to see this is to consider another core contrast – between our conceptions of magic and metaphysical free will, the examples mentioned above, and our conceptions of phenomenal experiences. It is easy for me to be convinced that no such thing as magic exists, or, indeed, is even possible. If I analyze what I have in mind by "magic", something like "a mechanism that isn't really a mechanism" seems plausible, and it's easy to see why no such

thing could even exist. Similarly, when I think about libertarian metaphysical freedom of the will, I can be brought to see that no such phenomenon (at least in our world) could answer to it. What's more, it's not hard to see how I came by such a conception even if nothing corresponds to it. However, when it comes to our concepts of phenomenal properties – concepts that have what I called elsewhere "substantive and determinate contents" – it's very hard to see how we even came by the relevant concepts unless there were existent phenomena that gave rise to them. This is of course very quick, but I argued for this in more detail, specifically targeting Rey, in Levine (2001, chapter 5).

Let's turn to the anti-materialist position on the first-order problem and see how the meta-problem plays out for that side. Of course there isn't just one anti-materialist position. One could be a substance dualist, property dualist, strong emergentist, and even panpsychist (though sometimes this is seen as a more liberal version of physicalism). I am not going to survey the various forms of dualism and see how the meta-problem arises for each of them. Rather, I will sketch a position I strongly incline toward, a form of strong emergentism, and see how the meta-problem plays out for this view.

While the terminology in this area is not standardized, by any means, what I mean by "strong emergentism" is this. Phenomenally conscious states and events are strongly emergent with respect to the underlying physical-functional phenomena of the brain, or central nervous system, just in case they aren't realized in the underlying phenomena. One way to put it is that phenomenally conscious states and events do not supervene on the underlying physical-functional states. There is a metaphysically possible world in which the same physical-functional events take place, governed by the same purely physical laws, and yet either there are no phenomenally conscious events or they differ from those of this world. If this is the situation, then I say that the phenomenally conscious phenomena are basic, and in addition to the fundamental entities, properties, and laws that govern the purely physical (or non-mental) world, there are in addition basic mental properties and basic psycho-physical laws.

Terminology here gets tricky. I am using the word "basic" to contrast with "realized" (or "supervenes on"), not with "complex" or "composed", which is why I steered away from the term "fundamental". Sometimes people hear the description of a property or a law as "fundamental" to mean that it applies only to so-called fundamental entities, the objects that are the ultimate building blocks of all other objects, like elementary particles, or fields. While those properties and laws are indeed basic in my sense, they are not the only ones. It is precisely the point of emergentism to posit properties and laws that apply at more macro levels of aggregation – including entire nervous systems – and yet are not realized in, or supervenient on the properties and laws governing the parts. It is clear from the panpsychist literature that rejection of the existence of (strong) emergence of this sort is a large part of what motivates the position. (See Bruntrup & Jaskolla 2017). Otherwise, why be saddled with the commitment to phenomenal properties at the lowest levels of nature, with its concomitant "combination problem"? But their argument is that it makes even less sense to posit strong emergence in nature than it does to attribute phenomenal properties to elementary particles.

So the position toward which I strongly incline is a strong emergentist one. Phenomenal consciousness is basic, not realized in underlying physical-functional mechanisms. To use the standard Kripkean metaphor, when God created the physical world, that wasn't in itself sufficient to bring consciousness into the world; rather, a new creative act was necessary. But of course the relation between the underlying physical-functional architecture and consciousness can't just be arbitrary either. There are obvious regularities that govern our conscious life and many of them involve both physical and mental properties. It's obvious that physical events and mental events (including phenomenally conscious ones) are tightly correlated. If you stick a pin in someone's arm they feel pain; if you remove someone's visual cortex they no longer see; and when you have conscious experiences you can describe them to someone, which entails making your body move in various ways. Whether these correlations involve genuine causal relations or not is perhaps contested, but that the correlations do exist is undeniable.

So in addition to the claim that conscious experience is strongly emergent we also have to posit some basic, and emergent, psycho-physical laws. That is, whether or not certain conscious phenomena emerge in certain physical systems is governed by laws relating the two domains, and these laws are basic in that they do not supervene on purely physical laws. On this view, then, while zombies are metaphysically possible because there are possible worlds that are "minimal physical duplicates" of this world without conscious creatures, they are not nomologically possible if we include the basic psycho-physical laws in the determination of what's nomologically possible. A minimal physical duplicate of our world would share all of its physical properties and purely physical laws, but lack the extra, basic, emergent psycho-physical laws that govern our world.

Accommodating psycho-physical correlations isn't quite enough, though. Among the materialist arguments that I find most compelling is the appeal to the myriad ways in which what I will call the "fine-tuned structure" of our conscious experience can be explained to a very large extent by the functional profile of the underlying physical mechanisms. For instance, take color experience. Leaving aside the explanatory gap that attends there being any experience at all, or one's color experience having the particular qualities that it does, there is clearly a lot about the structure of that experience that is explicable by appeal to underlying physical-functional mechanisms. On the opponent process theory, for instance, we can explain a lot about the similarity relations among color experiences along the three dimensions of hue, brightness, and saturation. Detailed investigation of the computational structure and implementing neural mechanisms starting from the retina through the visual cortex explains much of this structure. It's reasonable, then, to suppose that whatever psycho-physical emergent laws there are possess a unified structure that makes sense of this relation between physical-functional architecture and the fine-tuned structure of experience.

Here is how I speculate an emergentist position might accommodate this relation between the physical-functional, or computational architecture and the structure of conscious experience. First, treat conscious experience as the holding of a basic, intentional relation of acquaintance between the conscious subject and a virtual world of objects and properties. In a

sense I would endorse the almost universally deplored "Cartesian theater" model of experience (see Levine, 2018, chapter 12). What it is to have conscious experience, on this view, is just to stand in a primitive, or basic acquaintance relation to the objects of experience. What determines the nature of this experienced, virtual world are two factors: first, the precise computational description of the world produced by our perceptual and cognitive systems as a result of their interactions with the environment, and second, the nature of the basic psycho-physical law that translates this "script" provided by our cognitive and perceptual systems into the depictions of the world we experience phenomenally. So on this view there is no explanation – it's a brute fact – why objects whose surfaces possess a certain spectral reflectance profile are experienced as phenomenal reddish or greenish, as opposed to the reverse, or no way at all. However, the overall structure of the quality space of color experience would be explained by the nature of the visual representational structure that produces this experience. The "production" in question, then, would be a result of the basic psycho-physical law(s) that emerge(s) at that level.

Of course this is all too brief and sketchy. In particular, there is the obvious question of what is meant by a "virtual" object. The reason I need such a notion, and don't just say that the basic acquaintance relation holds between the subject of experience and the concrete physical objects with which she interacts in the world is that I want to cover cases of hallucination. When I'm hallucinating a pink elephant, I want to say that I'm acquainted with the pink elephant, but not that there is an actual pink elephant occupying some region of space-time with which I'm acquainted. What I have in mind, then, is something metaphysically in between, as it were, a genuine, concrete material object, on the one hand, and a merely intentional object on the other.

Philosophers have appealed to merely intentional objects when characterizing mental states like searching for the Fountain of Youth, or believing in Santa Claus. When asked, but what is it one is searching for or believing in, the answer is that these are merely intentional objects, not real ones. Of course just how to interpret the notion of a mere intentional object is a matter of dispute, but I think of it as merely a way of characterizing the state in question as purporting to refer to something when it really doesn't. However, I think there is something going on when we experience hallucinatory objects that is, to put it crudely, more real than what is going on when we merely think about something that fails to exist. There is a sense in which the pinkness is really in the experience, but not really in objective space-time.

The obvious model for what I have in mind is the old sense data theory. In a way, one could characterize my view as a version of sense data theory. Sense data are supposed to be items the mind perceives that do not occupy physical space-time but rather a mental realm. That is pretty much the metaphysical status I have in mind for virtual objects and their properties. The reason I don't use the term "sense data" – aside from trying to avoid the negative reactions that theory immediately provokes – is that sense data, as traditionally conceived, are just sensory qualities: colors, sounds, shapes, smells, and the like. However, what I have in mind is an organized world of objects and properties, a "Cartesian cinematic realization" of the combined perceptual and cognitive characterization of the world provided by

the mind as a whole. So there are books, tables, trees, animals that possess all sorts of features, many of which are the sensory qualities captured by the notion of sense data, but not limited to them.

So the brief sketch above is the candidate anti-materialist view I want to consider in light of the meta-problem. Clearly one issue for such a position, and the issue on which much support for materialism rests, is that emergentism seems to entail epiphenomenalism, at least if you add some plausible principle about the causal closure of the micro-physical. While many take this entailment to be a deal breaker – I certainly did in the past – others are willing to swallow it, if perhaps quite reluctantly. But related to this epiphenomenalism is a version of the meta-problem, and that is more serious.

Among the arguments briefly presented earlier for positing some sort of primitive, non-physically realized acquaintance relation has to do with the cognitive immediacy that attends our knowledge of our own experiential properties. My criticism of the phenomenal concepts strategy was that the only plausible account that would actually explain why entertaining thoughts about our experience through phenomenal concepts would yield cognitive resistance to materialism required positing some acquaintance-like relation for which a materialist account is lacking. The idea, then, is that it must be acquaintance with experience itself that is responsible for our coming to know (assuming for the moment it's true) that consciousness consists of the acquaintance relation.

The problem is, given the model described, it's hard to see how acquaintance could be the source of our knowledge of acquaintance. Suppose, as I described above, that our cognitive systems (including perception) work on physical-functional principles to realize computations that produce a comprehensive description of the world around us. The emergent, basic psycho-physical law then translates that description into the "movie" of conscious experience, populating a world of objects and properties with which we are acquainted. All judgments we make are, on this model, produced by the underlying cognitive system, and then they contribute to the constitution of experience. It is mysterious, then, how the nature of that experience could then itself become an input to our cognitive mechanisms and a source of the knowledge we seem to have about the nature of our experience.

In a way this problem is closely connected to the problem that Balog (1999) brings up for the conceivability argument. She argues that on the zombie hypothesis a zombie would make the same judgments about the conceivability of zombies that we would, yet for the zombie this is not possible, since she possesses no mental features not possessed by any other zombie. So if the conceivability argument fails for zombies, and we're in the same epistemic position they are, then it should fail for us. One way of replying is to reject the two assumptions of Balog's argument that first, the content of a zombie thought "about" phenomenal consciousness is really the same as ours, and second, the correlative claim that the basis of our justification for holding a belief about our phenomenal states is the same as it is for the zombie. This second claim, in particular, has been a staple of functionalist criticisms of anti-materialist views of qualia since Shoemaker's seminal paper "Functionalism and Qualia"

(Shoemaker 1984, chapter 9). If we posit an acquaintance relation that holds in us but not in the zombie, then we have a basis for distinguishing both the contents of our phenomenal judgments and their justification from those of a zombie.

However, if one appeals to acquaintance to sidestep the Balog argument, then we need to have some idea how conscious acquaintance can become a basis for judgment – both epistemically and causally. This seems to entail that the Cartesian theater of conscious experience cannot merely be an output – a byproduct? – of the physically realized cognitive machine, but also an input to it. How this is possible is a puzzle I still struggle with.

I want to end with some speculative thoughts about how one might address the problem just raised. One idea is to loosen the connection that is usually assumed to hold between identity and necessity. In particular, one might challenge the claim that if two properties/relations are distinct then there must be a possible world that in which their extensions differ. Notice that to the extent a lot of the debate over materialist identity theories (whether of the central state variety or functionalist) revolves around what's possible, this concerns only the entailment from identity to necessary connection. So one way to definitely challenge an identity thesis is to show that it is possible to have one side without the other. But it seems to me at least not obvious that one couldn't maintain that even if there is a necessary connection between the two properties/relations in question they might still be distinct.

If it were possible to occupy the very same cognitive state regarding acquaintance that we do now without phenomenal experience, then it is very difficult to see how such cognitive states could constitute genuine knowledge, especially of the kind we are supposed to have when it comes to experience. But if there is a kind of inherent necessary connection between being the kind of cognitive system we are and possession of the requisite experience, then the problem seems much less serious. What kind of metaphysics this would involve is not clear. The idea would be that there is something in the nature of intentional "scripts" of the kind we instantiate that necessitates an experiential reflection of it - a "script" that is necessarily produced as a movie. In a way, this might be a form of "pan-proto-psychism" as applied to certain informational states.

While allowing necessary connections between distinct existences (rejecting Humeanism) might help the problem, I don't think it's sufficient. We still need a way of making the cognitive immediacy of experience explicable in the nature of the relation between the cognitive states about acquaintance and the phenomenon of acquaintance itself. One possible line of investigation is to employ the notion of cognitive phenomenology. After all, it is when one is occurrently entertaining thoughts about one's experience that one gains knowledge of this acquaintance relation. So if there is an account of the cognitive phenomenology involved in the very thoughts about experience themselves that can ground the epistemic status of such thoughts, then perhaps we can overcome this problem entirely. At any rate, these are routes I intend to explore in future work.

References

Balog, K. (1999). "Conceivability, Possibility, and the Mind-Body Problem," The Philosophical Review 108: 497-528.

Bruntrup, G. & Jaskolla, L. eds. (2017). Panpsychism: Contemporary Perspectives. New York: Oxford University Press.

Kammerer, F. (2018). "Can you believe it ? Illusionism and the illusion meta-problem". Philosophical Psychology 31(1), 44-67

Levine, J. (2001). Purple Haze: The Puzzle of Consciousness. New York: Oxford University Press.

Levine, J. (2018). Quality and Content: Essays on Consciousness, Representation, and Modality. New York: Oxford University Press.

Rey, G. (1995). "Toward a Projectivist Account of Conscious Experience", in Metzinger, T., ed., Conscious Experience. Paderborn: Ferdinand Schöningh/ Imprint Academic.

Shoemaker, S. (1984). Identity, Cause, and Mind. Cambridge: Cambridge University Press.