

Asger Kirkeby-Hinrup

## *Two Caveats to the Meta-Problem Challenge*

**Abstract:** *I present two caveats to the meta-problem challenge to theories of consciousness. Chalmers suggests that a theory of consciousness that solves the hard problem should also inform us about the meta-problem, and vice versa. The first caveat is the view that mechanism M, the mechanism through which content becomes conscious, may be neutral with respect to the content it renders conscious. This means that there can be no systematic connection between M and conscious content. The second caveat concerns how we should treat the problem intuitions fuelling the meta-problem. I argue that we should award them no special status with respect to their explanatory power in relation to the hard problem.*

### **1. Introduction**

‘The Meta-Problem of Consciousness’ (Chalmers, 2018) is an enjoyable read and a fitting follow-up to Chalmers’ seminal paper on the hard problem of consciousness (Chalmers, 1995). Solving the meta-problem of consciousness roughly requires explaining our ‘problem intuitions’ about the hard problem. Chalmers suggests that the hard problem and the meta-problem are closely connected. The suggestion is that if we have a solution to the hard problem it should shed some light on the meta-problem, and vice versa (Chalmers, 2018, p. 35). Below, I will offer two worries I have about the proposed connection between the hard problem and the meta-problem. I will frame these worries in relation to what Chalmers calls the meta-

Correspondence:

Asger Kirkeby-Hinrup, Center for Functionally Integrative Neuroscience, Århus University, Denmark; Department of Philosophy and Cognitive Science, Lund University, Sweden. Phone: (+45) 30 70 49 77.

Email: [asger.kirkeby-hinrup@fil.lu.se](mailto:asger.kirkeby-hinrup@fil.lu.se)

problem challenge (MPC) for theories of consciousness. The MPC for theories of consciousness suggests that a theory of consciousness should not only explain mechanism M (the mechanism, presumably in the brain, that produces consciousness), but also inform us about the meta-problem. The MPC rests on the idea that, because the hard problem and the meta-problem are closely connected, a theory of consciousness that solves the hard problem should also make significant headway with respect to solving the meta-problem.

There are two distinct claims involved in the idea that the hard problem and the meta-problem are closely connected. The first is that a solution to the hard problem should shed light on the meta-problem. On this view, if we understand the nature of consciousness this will shed light on why we think (thought) consciousness is (was) hard to explain. Since the explanatory direction here goes from an explanation of consciousness to information about our problem intuitions, call this claim ECPI. The second claim reverses the explanatory direction and holds that if we understand our problem intuitions this will shed light on the nature of consciousness. With respect to this claim, the explanatory direction goes from an explanation of our problem intuitions to consciousness, let us call this claim EPIC.

The MPC for theories of consciousness is derived from ECPI and will be my main concern below in the first caveat. The second caveat concerns EPIC, but is related to the first.

## **2. First Caveat: Mechanism M May Be Content-Neutral**

The underlying idea in ECPI is that whatever explains consciousness should also help explain our judgment about consciousness. Chalmers writes:

...suppose there is a correct... theory of consciousness according to which mechanism M is the basis of consciousness. Then we can reasonably expect that mechanism M plays a central role in explaining our judgments about consciousness. (Chalmers, 2018, p. 36)

Chalmers thinks that such an expectation is reasonable because consciousness and judgments about consciousness are closely connected. However, this way of viewing things seems only partly right. It is certainly true that if a mechanism M is the basis for consciousness then whatever content X, such as a judgment, enters consciousness will do so at least in part because of M. What is not necessarily true is that M in any way influences the nature of X. To elaborate, it is

possible (and I think likely) that M is entirely impotent with respect to the contents of X. To use an analogy, think of consciousness as a television.<sup>1</sup> The television merely displays whatever information is fed through the appropriate circuits (mechanism M). The television may even display other televisions, or the internal physical components that make up televisions along with diagrams and explanations of how televisions work. It may display televisions from the 1950s or from the twenty-first century, or it may display broken televisions or faulty diagrams that purport to show how televisions work. The point is the television has no influence over the content it displays, the way it displays it, or any attitude toward whether the content is correct or not.

Importantly, I am not implying that the X cannot be modulated by cognitive processes or states. Indeed, there is an abundance of evidence from experimental psychology, cognitive science, and neuroscience that shows how X can be altered by manipulations to sensory stimuli, introduction of chemical agents, neural stimulation, and so forth. However, the fact that there are different ways to manipulate X is not evidence against the idea that M is content-neutral. It is possible that the manipulations occur upstream from M. On this view, when content reaches M it is faithfully rendered conscious with no further alterations. One objection to this idea would be to point to experimental data that show how a conscious percept can be modulated through continuous experimental manipulation. One example of this kind of modulation could be the gradual change blindness paradigm (see e.g. Grimes, 1996). But, even in this case, the changes to content may reflect processing upstream from M, and that when the updated visual stimulus reaches M it is faithfully represented with no alteration.

Now it may be objected that I am here misunderstanding Chalmers' assertions. To ascertain whether this is the case, a step by step look at the way he frames the MPC is helpful. Chalmers says:

For a realist, judgments about consciousness systematically reflect the character of consciousness. (Chalmers, 2018, p. 36)

By 'realist' Chalmers is clear that he means someone who thinks consciousness is real. However, it is unclear what 'the character of consciousness' is intended to capture. To my mind, there are two possible

---

<sup>1</sup> Please ignore the obvious connotations this analogy has to a Cartesian theatre of the mind.

interpretations. On the first (and I think most natural) interpretation, the *character of consciousness* means how consciousness appears to us. That is, consciousness from the first-person perspective. On this interpretation it certainly seems reasonable that how consciousness appears to us from the first-person perspective is what is driving many of the problem intuitions. However, if we assume that M is content-neutral in the sense described above, there is no connection between how consciousness appears to us and M itself. How consciousness appears to us will not reliably inform us about M. On the second interpretation, by the *character of consciousness* Chalmers actually means mechanism M. This interpretation fares worse than the first. In this case it is false that the character of consciousness *systematically reflects* M, since X can be a belief in any theory of consciousness. Chalmers says next:

It would be extremely strange if the mechanisms responsible for consciousness played no role, or only a minimal role, in generating judgments about consciousness. (*ibid.*)

Here it seems that Chalmers is talking specifically about M, albeit using the plural ‘mechanisms’ rather than the singular ‘mechanism’ used earlier. Now, on the view of M as content-neutral, it is not obvious that M plays a role in *generating* judgments about consciousness. Judgments are just content (X). M may just render conscious whatever X it received through the appropriate circuits.<sup>2</sup> I do not find this view extremely strange and would not be surprised if it turned out true. It is possible of course to interpret the word *generating* in a way that makes Chalmers’ formulation compatible with the view of M as content-neutral. For instance, if M’s rendering a subject conscious of X is what it means to *generate* a judgment about consciousness. But this interpretation does not sit well with the way Chalmers frames the meta-problem. Chalmers seems to suggest that M affects *what we judge about M*, i.e. the content of X. In fact, I think that this reading is the only way to preserve any *systematic reflection* of the nature of M in the judgment and make sense of the MPC. Chalmers continues:

---

<sup>2</sup> I take the formulation ‘through the appropriate circuits’ as neutral on the metaphysical relationship between consciousness and M. M and the upstream ‘appropriate circuits’ can be conceived of in both illusionist and realist terms. There is room for epiphenomenalists, or even property or substance dualists with interactionist leanings (if one allows that the top-down causal influences on the physical from the mental enter into the appropriate circuits somewhere upstream from M).

These mechanisms may not *wholly* explain our judgments. It is plausible that some further introspective processes beyond consciousness itself are required to generate the judgment. (*ibid.*)

While such introspective processes (e.g. thinking about how consciousness appears to me and how this appearance may be implemented in my brain) are likely to influence my judgments about consciousness and my inclinations to deem that there could be a problem, they will not improve the accuracy of my judgments about the nature of M. Again, on the view proposed above, introspective processes would be nothing more than additional content unable to reliably reflect the nature of M. So while introspective processes may shape the contents of my judgments, they are no help in establishing a systematic reflection of M. Finally, he says:

Still, a realist should expect that our judgments about consciousness are the way they are *because* consciousness is the way it is, or at least because the basis of consciousness is the way it is. So the mechanisms responsible for consciousness should play a central role in generating judgments about consciousness. (*ibid.*)

Here Chalmers seems to make the distinction I alluded to above. At least if you read *consciousness is the way it is* to mean how consciousness appears to us, and *basis of consciousness* to refer to M. If I am right that this is the distinction he wishes to make, I think he gets it backwards. It is very plausible that how consciousness appears to us influences our judgments about consciousness. However, the nature of M (possibly) wields no such influence.

What are the implications from this view on the MPC? If M is content-neutral the upshot seems to be that an understanding of M (i.e. solving the hard problem) will yield little to no information about the meta-problem. This, in turn, means that challenging theories of consciousness to say something about our problem intuitions is demanding too much. Of course, whether the view that M is content-neutral is available to a given theory of consciousness is likely to depend on how the theory in question is cashed out. *Prima facie*, it does seem like this view will be available to many of the mainstream theories of consciousness. For instance, higher-order thought theories that endorse the possibility of misrepresentation (e.g. Rosenthal, 2012) could easily adopt this stance. Similarly, integrated information theory (e.g. Tononi, 2005) can remain neutral about the relationship between (1) the mechanisms through which information is integrated (M), making X (a judgment about consciousness) conscious, and (2) the

content of X. I do not see any compelling reason to hold that (2) must systematically reflect (1). Similar lines of thinking seem available to many other theories.

In the above, I argued that mechanism M may be content-neutral. This provides ways of cashing out theories of consciousness to which the MPC does not apply. Now, some may object that I have fundamentally misunderstood what Chalmers proposes. Perhaps what is really entailed by the MPC is that a theory of consciousness (that solves the hard problem) needs to explain how consciousness appears to us the way it does.<sup>3</sup> After all, it is because consciousness appears to us the way it does that we have problem intuitions. I think this objection has some validity but does not save the MPC.

### **Second Caveat: Appearance, Reality, and Problem Intuitions**

In his presentation, Chalmers is clear that problem intuitions are not universal but submits that they are at least widely shared. This seems right. To many people consciousness is indeed puzzling. Chalmers then suggests that:

A fully adequate solution to the meta-problem should be able to explain not only why these intuitions are widely shared, if they are, but also why they are not universal, if indeed they are not. (Chalmers, 2018, p. 14)

A tall order, but here is an attempt to provide the requested explanations. With respect to the first, the need to explain why problem intuitions are widely shared, it is useful to look for the foundation upon which they rest. One plausible foundation is the fact that appearance and reality may come apart. We know this from our perception of the outside world: I may see a cow and think it is a horse. We know this from social contexts: I may think my wife is sad, but she is really just tired. We know this from introspection: I may think I am hungry, but really I am just procrastinating. And we know this from empirical sciences, where we may think we see a particle, but really it might as well be (or is also) a wave. The appearance–reality problem certainly seems to have a role to play in the hard problem, and consequently in our problem intuitions about it. Essentially, it

---

<sup>3</sup> This, after all, is one of the two central tasks of the hard problem, namely solving the ‘how’ conscious appears to us the way it does (the other is ‘why’).

applies because (to many people) how consciousness appears to us is seemingly at odds with what we think we know from natural sciences.

A further fact supports the idea that appearance and reality may come apart with respect to consciousness. This is the fact that people have different convictions about consciousness and its connection to physics. I do not purport to know the inner life of a sincerely convinced illusionist, but I presume that despite many similarities her experience of being conscious somehow differs from mine on certain points. At the level of judgments (or philosophical theory) it is even clearer that there is a chasm between conceptions of consciousness and ways the world can be. To see this, consider the variety of theories and conceptions of consciousness on offer. Clearly, many of these positions are mutually exclusive, i.e. if theory T is true the world must be in a certain way, and this precludes theory Y from being true. Additionally, many of these theories (e.g. illusionism) hold that reality is not how it appears to us (at least with respect to consciousness). So again, the fact that there are differing conceptions of the relationship between consciousness and the world that cannot all be true further supports the notion that the appearance–reality problem plays a central role in the hard problem and consequently our problem intuitions. From this, one may want to refine the claim to something along the lines of: we have competing theories about how the world could be for things to appear as they do, but cannot determine which is correct. This claim seems true, but also leaves out the possibility that none of our theories are correct, so perhaps some would prefer the even stronger claim: we have no idea how the world must be for things to appear the way they do and how to determine whether we were right even if we had the right idea. I think this claim is too strong, given that we have at least *some* ideas. Be that as it may, demystifying the hard problem by seeing it in the light of the appearance–reality distinction goes some way to understand what fuels our problem intuitions.

Briefly, I will now turn to the second call for an explanation pertaining to the fact that problem intuitions are not universally shared. From the fact that problem intuitions are indeed not universal (as Chalmers seems to admit), I am at loss as to why we should treat them differently to any other intuition that is widely shared but not universal. For instance, many people around the world may share the intuition that there is a problem with the rising economic inequality across the globe. They may even share the intuition that this problem somehow turns on historical and institutional factors along with

distribution of political (and military) power. This intuition is widely shared, but not universal. It is hard to see how any science of international politics, distributions of wealth and power (or whatnot), let alone consciousness would *have* to account for this intuition. Similarly, many people share the intuition that there is a fundamental issue associated with the measurement problem in quantum physics. But we would not expect a theory of quantum physics that solved the measurement problem to have anything further to say about this intuition. That problem intuitions (in any domain) are not universal just shows that what I think (or intuit) presents a problem is ultimately arbitrary. If this is correct, this means that a theory of consciousness no more needs to explain problem intuitions about the hard problem than it needs to explain problem intuitions about the measurement problem. If our problem intuitions are treated like any other X by M, they are detached from the nature of the processes that make them appear to me and carry no evidential weight as to the nature of the problem they are about. The thought that consciousness is an illusion may occur to me even if consciousness is not an illusion. The opposite is also possible, i.e. the thought that consciousness is real may occur to me even if consciousness is not real. In the same vein, the intuition that there is a hard problem of consciousness may occur to me even if there is no hard problem of consciousness, and vice versa.

### References

- Chalmers, D.J. (1995) Facing up to the problem of consciousness, *Journal of Consciousness Studies*, **3** (1), pp. 200–219.
- Chalmers, D.J. (2018) The meta-problem of consciousness, *Journal of Consciousness Studies*, **25** (9–10), pp. 6–61.
- Grimes, J.A. (1996) On the failure to detect changes in scenes across saccades, in Akins, K. (ed.) *Perception*, Oxford: Oxford University Press.
- Rosenthal, D.M. (2012) Higher-order awareness, misrepresentation and function, *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **367**, pp. 1424–1438.
- Tononi, G. (2005) Consciousness, information integration, and the brain, in Steven, L. (ed.) *Progress in Brain Research*, Amsterdam: Elsevier.