

# Untangling the Hard Problem of Consciousness

Todd L. Duncan  
Center for Science Education, Portland State University &  
Department of Physics, Pacific University  
duncant@pdx.edu

## Abstract

Subjective experience is difficult to understand within the framework of the “physical universe,” which is a particular model that only incompletely captures the properties of the real universe. Since this physical model focuses on objective, functional properties of the world and omits the subjective aspects, it’s not surprising that we have difficulty describing subjective experience from within its framework. In this paper I explore this perspective as a way to shed light on the hard problem of consciousness and to help pave the way for a worldview that is consistent with scientific knowledge but also provides a clear home for subjective experience and human meaning.

## Introduction

The so-called “hard problem” of consciousness can be expressed as the challenge of explaining how it can be that *subjective experience* arises from *physical* processes in the brain (Chalmers 1995, 2002a, b). Why is it that certain physical brain states are accompanied by the experience that there is “something it is like” to be in those states, even though such subjective experience seems extraneous to the physical description, and even though most physical states are apparently not accompanied by such experiences (Nagel, 1974)? In this paper I suggest a way to untangle the difficulty by clarifying the origin of the problem. Specifically, I argue that the hard problem is a byproduct of the models we create to describe different aspects of the universe we experience. The aspects we label with the category “physical processes” or “physical reality” – including physical brain processes – constitute a *model* that captures some aspects of the deeper underlying reality very well, but ignores other aspects of reality. Similarly, those aspects we label as “subjective experience” or “consciousness” also provide only an incomplete description of the underlying reality. In a vague sense we are certainly aware that each category is only an incomplete model, but in an immediate sense it’s easy to forget this basic fact and find ourselves tangled in confusion as a result.

When we view the hard problem from this perspective, our earlier phrasing is revealed as misleading. Subjective experience does *not* arise (in a derivative sense) from physical processes in the brain. In fact it would make no sense if experience *did* arise from physical processes, because physical processes constitute a model of reality that specifically focuses on the functional aspects and filters out and *excludes* the aspects of reality that constitute conscious experience. Rather, subjective experience arises as part of the deeper substrate of reality that *also* produces the aspects we model and describe as “physical” processes in the brain. The incompatibility of our constructed categories need not imply an incompatibility in the underlying reality these categories were invented to

partially describe. The rest of this paper is devoted to clarifying and solidifying this essential point.<sup>1</sup>

## Models and Reality

One way to see that the problem is created by the particular categories we use to describe reality is simply to note that the problem *does not exist* without the categories. Imagine awakening to experience the world from a pre-scientific perspective. This must have been the case for many thoughtful humans prior to the last few hundred years, and even for many people today who are not absorbed in the scientific world view. From this perspective, the most obvious fact of your experience is the awareness that there is “something it is like” to be you. Subjective awareness is such a dominant feature of your experience with the world that there is no question it is real, and no problem of trying to “explain” it. It just is.

What happened to transform such an obvious feature of our experience into a puzzle that seems so intractable we’re willing to consider explaining it away and describing it as an illusion, not essential to reality at all? Vastly oversimplified, one might summarize the historical story as something like this: By making use of our consciousness, we learned to notice patterns in our surroundings, and codified these patterns as laws of nature to form a model that we now call “physical reality.” This model is tremendously powerful for explaining and organizing a vast range of our experience. As a result, it’s easy to become so absorbed in that *part* of the universe (the functional model which is the physical universe) that we start to equate the model to all of reality, and find ourselves alienated from the real universe we started with. This makes sense as a natural progression of our developing understanding (it was probably necessary to focus in this way in order to make progress). And in many situations, we can safely ignore the fact that our model is not identical with reality. But one obvious limitation to a model is that it does not contain the model-maker within it. So a puzzle like the hard problem of consciousness is just the sort of situation where we *cannot* safely ignore the limitations of a model—particularly a model specifically designed to focus on the objective, functional aspects of reality and ignore those subjective aspects related to experience! Why should we be surprised that we have difficulty finding a place for conscious experience within a model essentially *designed* to leave out conscious experience? From this perspective, the puzzle is a reminder of the limitations of even our most successful models, and pushes us to take the next steps in expanding the models to be more inclusive of the aspects of our experience that have been omitted so far. Following the insight of Kenneth Craik, we need to move toward extending our theory to be as wide as reality and guard against letting our “perception of reality become as narrow as [our] theory.” (Craik, 1943)

The situation is captured well by Alexandre Koyre’s description of the Newtonian synthesis in the history of physics:

I have been saying that modern science broke down the barriers that separated the heavens and the earth, and thus it unified the universe. And this is true. But, as I have

---

<sup>1</sup> Harrison (2003) provides an excellent articulation of the general perspective within which my suggestion here is embedded.

said, too, it did this by substituting for our world of quality and sense perception, the world in which we live, and love, and die, another world—a world of quantity, of reified geometry, a world in which, though there is a place for everything, there is no place for man. Thus the world of science—the real world—became estranged and utterly divorced from the world of life, which science has been unable to explain—not even to explain away by calling it ‘subjective’. This is the tragedy of the modern mind which ‘solved the riddle of the universe’ but only to replace it by another riddle: the riddle of itself. (Koyre, 1967)

The conclusion is that the hard problem of explaining conscious experience is largely a puzzle of our own creation. Pointing this out does not make the problem any less real. Rather, it serves to highlight that the origin of the problem lies within our process of model-making, and directs us toward better understanding this process and the limitations of any particular model in order to seek a solution.

### **An Analogy to Highlight the Essence of the Problem**

A loose analogy may help clarify the essence of my point. The analogy is simplistic to the point of perhaps seeming ridiculous. But such oversimplification serves to make the origin of the paradox clearly visible in the much more subtle case of the hard problem of consciousness.

Imagine watching a busy city street from the balcony of a high rise apartment. More is happening than you could ever hold in your awareness all at once, but as you watch, patterns begin to emerge: the flow of people and cars, the structure of the streets and buildings, etc. With a pencil and sketch pad you could capture some of this order by drawing black and white stick figure “snapshots” of the scene (for example, locations of people) at different times, say at thirty minute intervals. This set of sketches would constitute a model that captured certain features of the happenings in the city: locations and relationships and a sense of the movement. But clearly the sketches also omit some features of the scene, such as color, sound, and what happens in between snapshots.

Now imagine getting immersed in thinking about your *model* of the city street—the black and white sketches. You could refine the sketches to show more detail, and learn to make predictions (e.g. every day at 5 p.m. a flood of people emerges from a particular building). You might get lost in the power of the model to the point that you took it to *be* reality—thinking that your sketches in some sense capture everything real there is to know about the city street.

But this situation would leave you with a nagging problem visible upon deeper reflection—a problem we might refer to as “the hard problem of color.” Namely, even though your model would explain or predict most features of the city street you could articulate (when people emerge from buildings, the flow created by traffic lights, etc.) it would not explain this “extra” experience of the real world, namely, seeing color. As with a reductionist response to the hard problem of consciousness, you might be able to explain this problem away. After all, is color really necessary to explain anything that you would seek to explain with your model? Yet something still seems wrong—it’s such an obvious fact of real experience that seems to have no analog in the model.

Of course, from an outside perspective in this analogy, the solution is obvious. Namely, color *is* a part of the underlying reality that is omitted from our model, and will never emerge from our model no matter how we contort it. I contend that the hard problem of conscious can be resolved in a similar manner.

The analogy helps clarify that the question, “How can subjective experience arise from a physical brain process?” is really improperly stated. It’s loosely analogous to asking, “How does color arise from black and white pencil sketches?” There we can clearly see that the point is the color does *not* arise from the black and white pencil sketches (even though the pencil sketches and the color aspects of the scene both arise from the same real, underlying scene itself, so in this sense they are linked). Similarly, what we should be noting more accurately about brain processes and conscious experience is that the same underlying unfolding of reality that gives rise to the phenomena we model as physical brain processes, *also* give rise to the subjective experience of consciousness.

### **How Does this Fit within Physics?**

One might argue that progress in physics has made this whole discussion irrelevant. After all, if a microscopic physical description of the world is causally complete, then there is no room for anything more. It’s all well and good to say that there is “more to the world” than can be described by our physical model. But if in fact everything we observe *can* be accurately described by the physical model, then practically speaking there is *not* more to the world.<sup>2</sup>

Our color analogy helps reveal a sleight of hand at the heart of this argument. If one is allowed to filter what constitutes an observation of reality to include only the functional properties, then of course one will find nothing that cannot be explained by a model which is designed to describe only the functional properties. (I assume here for simplicity that the model is complete within its domain of applicability.) This is precisely the danger Craik (1943) warned of in commenting that “instead of [our] theory being as wide as reality, [our] perception of reality may become as narrow as [our] theory.” In the analogy from the previous section: if we are allowed to rule out “color” as a true observable part of reality (perhaps labeling it an “illusion” in some way), then we can convince ourselves that our black and white sketches capture all there is to know.

What’s more, even within physics it’s becoming clear that there is more going on than is captured by a naive mental image of the physical world (so-called “naive realism” or more precisely, “Cartesian representational realism”). This is perhaps most apparent in the foundations of quantum mechanics, where experimental results force upon us the realization that the rules and restrictions on observable states can be stored in what might be called more flexible ways than we generally assume when naively picturing physical

---

<sup>2</sup> It is important to note (see e.g. Anderson, 1972 and Leggett, 1987, p. 116) that just because the behavior of macroscopic systems is *consistent* with microscopic laws of physics, this does not necessarily imply the ability to *construct* or deduce macroscopic behavior purely from the microscopic laws, with no other input.

reality.<sup>3</sup> An analogy serves to illustrate the impact of quantum mechanics on our mental model of an underlying physical reality.

Picture a physical object such as an arrow that has a definite orientation in space. Implied by this, and built into your picture, is the assumption that all three components of the length of the arrow (in the  $x$ ,  $y$ , and  $z$  directions, say) have definite values that exist in reality, whether or not they are measured. The fact of the existence of the arrow as a physical object in Euclidian 3-dimensional space demands a relationship between these cartesian components of this physical arrow; namely,  $L^2 = x^2 + y^2 + z^2$ , where  $L$  is the length of the arrow and  $x$ ,  $y$ , and  $z$  are the three components of its length. One of the most profound insights from quantum mechanics can be expressed in this analogy as the realization that at a fundamental level the  $x$ ,  $y$ , and  $z$  components of the arrow's length do not all have well defined values at the same time. This fact allows a more flexible relationship between the components which is revealed in experiments such as the tests of Bell's inequality and similar theorems (see e.g. Gröblacher et al., 2007).

Experts in quantum foundations will note that there are many options for how to interpret the experimental results, and my analogy in the previous paragraph chooses one such interpretation. But the experimental results stand alone, regardless of interpretation, in confronting us with a simple and important fact. Correlations exist between observable properties of systems that cannot be explained by imagining that the correlations are stored as properties of a local, physical reality as we normally picture it. It's as if nature is trying to remind us not to take the physical model too seriously. This point is nicely emphasized by Fuchs (2002), "Try to give quantum mechanics a [Cartesian representational] realist interpretation – you can do it...[but] you find yourself contorting yourself beyond belief. It's as if nature is telling us for the first time, 'Please don't interpret me in a [Cartesian representational] realist fashion. I can't stop you, but please don't.'"

Here's why this discussion of physics is relevant. What makes the hard problem hard is the implicit assumption we make in asking how conscious experience can arise from physical processes. What we're really getting at with this question is, "How can this subtle thing we call conscious experience arise from what we know are *really* physical processes?" The whole basis for the hard problem of consciousness results from thinking that reality *is* essentially what we describe by physical processes which unfold as required by their underlying physical reality, and there is nothing else needed beyond that. We seem to have reality pegged simply by functionally describing the unfolding of the physical processes—except for this pesky thing called conscious experience that nags at us because it doesn't quite fit the framework. From that perspective, consciousness seems so insignificant in the grand scheme of things—for all we know it only occurs here on this tiny planet in all the vast ocean of the universe, the rest of which, we may assume, can be completely captured by the physical model. Why invent anything beyond this, and mess with our tidy scheme, just to incorporate such a fluke occurrence as consciousness, that we can easily explain away as irrelevant?

---

<sup>3</sup> See von Baeyer, 2003 pp. 163-182 for a conceptual introduction to these insights of quantum mechanics. More technical details can be found, for example, in the discussion of Bell's inequality by Greenstein and Zajonc, 2003 and by Mermin, 1985, and discussions of Bell's inequality and the Kochen-Specker theorem by Peres, 1993.

The importance of the insights from foundations of physics in this context is to point out that we actually need more than just the physical model to completely describe reality *anyway*, for reasons *other* than trying to explain subjective experience. Quantum mechanics practically demands that we recognize this, but it's also apparent if we look deeply even at classical physics. For example, something more to reality is needed as a means of storing the laws that tell our physical particles what properties to have and what laws to follow. And for any model, an aspect of reality is also needed to house the model-maker. The "capacity to create such a model" (some aspect of our consciousness, if you like) is a property of the universe that is not contained within the model itself.

The conclusion here is so important that it bears repeating: The simple mental image of local, physical reality that we hold in mind when expressing the puzzle, "How can physical processes give rise to subjective experience?" is simply *not* an accurate picture of reality. Physics itself points to the need for a deeper substrate of reality in which even what we call "physical" is embedded or represented; a substrate demanded not just by the need to explain conscious experience, but also by the need to explain quantum phenomena, the ability to encode properties of systems and the laws of nature, and even the existence of models and the capacity to create models. More speculatively, there is an emerging point of view that might combine several of these types of insights from physics into an information-based description—treating information as a fundamental property of the universe on par with energy or mass or charge, etc. (see e.g. Zeilinger 1999, von Baeyer 2003, Duncan and Semura 2007). It's not obvious that such a model will allow us to better explain subjective experience, and it certainly does not automatically show us how to do so. But by trying to include information as a fundamental aspect of reality, it serves as one example of a perspective that might help expand our model of reality beyond the one that creates the hard problem of consciousness.

## Conclusion

The hard problem of consciousness is a puzzle created by taking too seriously a powerful but limited model of reality. Most of the time it's valuable as a scientist to be completely immersed in a model, as if it were reality, to follow its leads and draw new conclusions and predictions. It's useful to take a model very seriously and see how far it will take us.

But when we encounter a paradox that is *caused* specifically by our unjustifiably *equating* a model with reality, we need to step back and remember the limitations of the model. A clear awareness that the hard problem is of this sort helps get our thinking unstuck. We won't be able to solve the problem from within that model, any more than we can find color in black and white sketches. Once we recognize this, the way is opened for an expansion of our models and a deeper glimpse of the underlying reality that any model describes only incompletely.

## Acknowledgments

I'm grateful to James Butler, Amanda Duncan, Chris Fuchs, Doug McCarty, and Katie Whalen for valuable discussions.

## References

- Anderson, P.W. (1972), 'More is different', *Science* **177** (4047), pp. 393-396.
- Chalmers, David J. (1995), 'Facing up to the problem of consciousness', *Journal of Consciousness Studies* **2** (3), 1995, pp. 200-219.
- Chalmers, David J. (2002a), 'The puzzle of conscious experience', *Scientific American (The Hidden Mind)*, pp. 90-100.
- Chalmers, David J. (2002b), 'Consciousness and its place in nature', in D. Chalmers, ed., *Philosophy of Mind: Classical and Contemporary Readings* (Oxford: Oxford University Press).
- Craik, Kenneth (1943), *The Nature of Explanation* (Cambridge: Cambridge University Press).
- Duncan, Todd and Jack Semura (2007), 'Information loss as a foundational principle for the second law of thermodynamics', *Foundations of Physics* **37**, pp. 1767-1773.
- Fuchs, Chris (2002), 'The anti-Växjö interpretation of quantum mechanics', arXiv:quant-ph/0204146v1.
- Greenstein, George and Arthur G. Zajonc (1997), *The Quantum Challenge: Modern research on the foundations of quantum mechanics* (Boston and London and Singapore: Jones and Bartlett Publishers).
- Gröblacher, Simon et al. (2007), 'An Experimental Test of Non-local Realism', *Nature* **446**, pp. 871-875.
- Harrison, Edward (2003), *Masks of the Universe* (Cambridge: Cambridge University Press).
- Koyre, Alexandre (1967), *Newtonian Studies* (Chicago: University of Chicago Press).
- Leggett, A.J. (1987), *The Problems of Physics* (Oxford and New York: Oxford University Press).
- Mermin, N. David (1985), 'Is the moon there when nobody looks? Reality and the Quantum Theory', *Physics Today* **38**, pp. 38-47.
- Nagel, Thomas (1974), 'What is it like to be a bat?' *The Philosophical Review* **83** (4), pp. 435-450.
- Peres, Asher (1993) *Quantum Theory: Concepts and Methods* (Dordrecht and Boston and London: Kluwer Academic Publishers).
- von Baeyer, Hans Christian (2003), *Information: The New Language of Science* (Cambridge, MA: Harvard University Press).
- Zeilinger, Anton (1999), 'A foundational principle for quantum mechanics', *Foundations of Physics* **29**, pp. 631-643.