

Bertrand Russell and the Problem of Consciousness

The Problem of Consciousness

People often talk about consciousness as a mystery. But there isn't anything mysterious about consciousness itself; nothing is better known to us than our own experiences. The problem of consciousness arises when we try to work out how the reality of my consciousness fits in with the material reality of my body and brain. If you cut open the top of my head and peer inside you'll find extremely complex physical processes ultimately constituted of neural firings. On the face of it, however, what you won't find – no matter how much you prod around inside my skull – are feelings, emotions, or experiences. The reality of my conscious experience as it is known to me through introspection seems to have nothing in common with the reality of my brain as it is known to empirical science.

Dualists take this apparent breach in nature at face value, holding that consciousness is a property of the immaterial soul; mind and matter causally interact with each other, but in themselves they are entirely different kinds of thing. However, the more we learn about the brain the harder dualism is to sustain, as detailed investigation of the brain reveals no sign of the action of a non-physical mind.¹ It is thus desirable to try to overcome appearances, and to make sense of the materialist thesis that conscious states just are, or are wholly constituted by, material processes in the brain.

The problem is that there are two powerful philosophical arguments which seem to show that materialism can't possibly be true: the *knowledge argument* and the *conceivability argument*. The knowledge argument begins with the claim that physical information about the body and brain, no matter how detailed and accurate, will always leave out one crucial bit of information about experiences: namely *what it's like to have them*. A colour blind neuroscientist cannot discover what it's like to see red by studying the neuroscience of colour experience. The only way you can find out what it's like to see red is by actually having a red experience. But if there's something about colour

¹ See Papineau 2001 for a detailed defence of this form of anti-dualist argument.

experience that no amount of brain science can teach us, it seems to follow that there is more to colour experience than the neurophysiological processes which are the subject matter of brain science.²

The contemporary form of the conceivability argument concerns *philosophical zombies*: perfect physical duplicates of living, functioning human beings which have no conscious experience. In contrast to a Hollywood zombie, a philosophical zombie walks and talks and behaves in every way just like a normal human being. And the reason a philosophical zombie behaves just like a human being is that the physical workings of its body and brain perfectly resemble that of an actual human being. And yet there is nothing that it's like to be a philosophical zombie. When you stick a knife in one it screams and runs away, but it doesn't actually *feel* pain. When crossing the road a zombie looks carefully both ways waiting for the traffic to clear, but it has no visual or auditory experience of the world around it. The lights are on but nobody's home.

Nobody thinks zombies are real, but many philosophers think they are conceptually coherent, in the sense that armchair reflection does not reveal any contradiction or incoherence in the idea of a zombie. Proponents of the conceivability argument infer from this that zombies are *possible*, that if the universe had gone very differently there could have been such things. The trouble for the materialist is that the *mere possibility* of zombies is inconsistent with materialism. If my consciousness is nothing over and above the material functioning of my body and brain, then you couldn't possibly duplicate the material functioning of my body and brain without thereby duplicating my conscious experience. But if zombies are even possible then you *could* in principle duplicate the material functioning of my body and brain without duplicating my conscious experience.³

We now face a deep dilemma. Science seems to show us that consciousness must be physical; philosophy seems to show us that consciousness cannot possibly be physical. Which side will blink first in this existential stand off? Fortunately there is a way forward, as Bertrand Russell saw nearly ninety years ago.

The Russellian solution

In his 1927 publication *The Analysis of Matter* Bertrand Russell pioneered a novel approach to bringing mind and matter into harmony. A similar view was expounded by Arthur Eddington in his Gifford

² Jackson (1982, 1986), Robinson (1982).

³ The term 'zombie' is due to Robert Kirk (1974), but the zombie problem was most influentially pressed by David Chalmers (1996/2008).

lectures of the same year.⁴ Sadly this new approach was swiftly forgotten about, perhaps because it did not fit with the physicalist zeitgeist of the latter half of the twentieth century. However, it is currently enjoying a revival, with a number of philosophers defending various forms of the Russell-inspired view which has become known as ‘Russellian monism.’⁵ Scholars debate how similar Russellian monism in its various forms is to the view Russell actually held in in the twenties.⁶ However, all Russellian monists begin from a certain startling epistemological thesis which Russell developed in *The Analysis of Matter*. In this section I will outline this thesis.

In the public imagination, physics is well on its way to giving us a complete account of the nature of space, time and matter. We’re not there yet: our best theory of the very big, i.e. general relativity, does not fit with our best theory of the very small, i.e. quantum mechanics. It is hoped nonetheless that one day these difficulties will be ironed out, and the public will be presented with the Grand Unified Theory which will tell us the complete fundamental nature of the material world.

However, as Russell recognized, there is an inherent difficulty in this understanding of physics, even before we get to the problem of consciousness. The problem is rooted in the fact that from Galileo onwards physics has worked with an extremely austere vocabulary, limiting itself to mathematical and causal predicates. It is not clear that such an austere vocabulary is capable of giving an adequate description of the nature of matter. It is natural to think that a mathematical models abstracts from the concrete reality of its subject matter. For example, a mathematical model in economics abstracts from the concrete reality of what is bought and sold, and the nature of labour. But physics is essentially offering us a mathematic model of fundamental physical entities such as electrons. Hence, in so far as we think a mathematical model abstracts from concrete reality, we ought to hold that physics abstracts from the concrete reality of an electron. As Russell put it ‘Physics is mathematical not because we know so much about the physical world, but because we know so little.’⁷

This difficulty arising from the austerity of physical vocabulary is avoided if we have a correspondingly austere conception of physical reality. *Dispositional essentialists* believe that there is nothing more to possessing a physical property such as mass than being disposed to behave a certain way, in the case of

⁴ These lectures were published the following year (Eddington 1928).

⁵ See Alter & Nagasawa 2015 for recent collection of essays on the topic.

⁶ Contrary interpretations of Russell’s view are argued for in the papers Stubenberg and Wishon in Alter & Nagasawa 2015.

⁷ Russell 1927: 125.

mass resisting acceleration and attracting other things with mass.⁸ Things on this view are not so much *beings* as *doings*: if you understand what an electron *does* you know everything there is to know about its nature. Assuming dispositional essentialism, it is more plausible that physics can completely characterise the nature of physical entities; a mathematical model can capture what an electron does, and in doing so tells us what it is.

However, there are powerful arguments against the intelligibility of dispositional essentialism. Most discussed is the worry that attempts to characterise the nature of properties, under the assumption of dispositional essentialism, lead to vicious regress.⁹ For any given disposition X, we understand the nature of X only when we know what it's *manifestation*, i.e. the property it gives rise to when manifested. For example, the manifestation of flammability is burning; we only know what flammability is when we know that burning is its manifestation. However, assuming dispositional essentialism the manifestation of any disposition X will be another disposition, call it 'Y'. To know the nature of X we need to know the nature of Y. But we can only know the nature of Y by knowing the nature of *its* manifestation, which will be another disposition, call it 'Z'. To know the nature of Z we need to know the nature of its manifestation, and so on *ad infinitum*. The buck is continually passed, and hence an adequate understanding of the nature of any property is impossible, even for an omniscient being; in other words a dispositional essentialist world is unintelligible. Riffing off G. K. Chesterton's line that 'We cannot all live by taking in each other's washing,' Russell made the following observation:

There are many possible ways of turning some things hitherto regarded as 'real' into mere laws concerning the other things. Obviously there must be a limit to this process, or else all the things in the world will merely be each other's washing.¹⁰

What Russell had realised was that far from giving us a complete picture of the nature of matter, physics is telling us absolutely nothing of the intrinsic character of matter. Physics tells us what matter *does*, but it tells us nothing about how it *is* in and of itself.

A similar story is true as we move up the physical sciences. Neuroscience characterises brain states in terms of what they do – their causal role in the brain and in the production of behaviour – and in terms of the chemical nature of their constituents. Chemistry in turn characterises elements in terms of their causal role in the overall system mapped out by chemists, and in terms of their physical constituents.

⁸ Bird 2007, Ellis 2001, 2002, Molnar 2003, Mumford 2004.

⁹ Campbell 1976, Robinson 1982, Blackburn 1980, Armstrong 1997, Heil 2003, Lowe 2006.

¹⁰ Russell 1927: 325.

And physics, as we have seen, characterises physical entities in terms of their behaviour. Throughout the whole hierarchy we find not a trace of intrinsic character. All that is revealed is what Russell called the ‘causal skeleton’ of the world; the physical sciences teach us nothing about the flesh and blood of reality.

Is there anything we do know of the intrinsic character of matter? According to Russell the only ‘window’ one has onto the nature of matter is provided by attention to one’s own mind. If dualism is false, and my mind is my brain, then the qualities that constitute my mind are one and the same as the qualities that constitute my brain. Those very qualities I find in attending to my experience – the feeling of anxiety, the sensation of an itch, the character of a red experience – constitute at least in part the nature of my brain. This line of reasoning led Russell to make the following shocking declaration ‘what a physiologist sees when he examines a brain is in the physiologist, not in the brain he is examining.’¹¹

How on earth could that be? As we began this article by observing, the material processes I observe in my brain seem to have nothing in common with the feelings, emotions and experiences I discover in my conscious mind. But thanks to Russell we can now see that this judgment depends on an overestimation of what the physical sciences are able to tell us about the nature of the physical. When we judge that brain states cannot possibly be feelings, we are assuming that we have some grip from the physical sciences of what a brain state is, of its intrinsic character. We feel we know what a brain state is (from physical science), and in knowing this we know that it cannot possibly be a feeling. But Russell has taught us that we know absolutely nothing of what a brain state is (from physical science), and hence know nothing (from physical science) that could exclude the possibility that brain states are feelings.

How does Russell’s insight help us with the knowledge argument? If red experiences are material states of the brain, why is it that their complete nature cannot be known from neuroscience (as is demonstrated by the fact that a colourblind scientist could never work out discern what it’s like to see red from studying brain science)? If Russell is right that brain science teaches us nothing about the intrinsic character of brain states then the problem disappears. Red experiences are states of the brain, but physical science can never teach you the intrinsic character of those states, which explains why neuroscience cannot teach you everything there is to know about those states. To learn about the intrinsic character of a brain state which is a red experience you have to actually have that brain state/red experience constitute part of your brain/mind.

¹¹ Russell 1927: 320.

What about zombies? The apparent conceivability of zombies arises from the fact that when we conceive of a zombie we are conceiving of its material nature in terms of its causal skeleton, i.e. in terms of its abstract mathematico-causal structure, rather than in terms of the intrinsic character that realizes that structure. If you could look into my head and somehow magically perceive the intrinsic character of my brain states, it would be apparent to you that my brain states are feelings, and a zombie version of me would cease to be conceivable. But so long as you are conceiving of my brain states in terms of what they do rather than what they are, it will remain concealed from you that my brain states are essentially feelings, and hence it will be possible for you to coherently suppose that my brain states could exist in the absence of consciousness.

The problem of consciousness arises, therefore, because we think the physical sciences tell us more than they do. Appreciating Russell's insight points to the way to a radically new theoretical approach to consciousness.

Turning the mind-body problem on its head

For most of the latter half of the twentieth century philosophers adopted what we might call a 'Brain First' approach to the problem of consciousness:

The Brain First approach: The physical sciences can in principle give us a complete understanding of the metaphysical reality inside a human head. Neuroscience informs us of the nature of brain processes; chemistry informs us of the nature of their chemical constituents; physics informs us of the nature of atomic and sub-atomic constituents of chemicals. The job of the philosopher of mind is not to add to this account of what's going on inside our skulls, but rather to try to squeeze consciousness into the story provided by the physical sciences. The aim of a philosophical theory of consciousness is to allow us to identify consciousness with something the physical scientist has already told us about.

What has bothered many about this approach is that it always seems to involve *redefining consciousness*. You start off focusing on, say, a feeling of pain, wanting to know where feelings of pain fit into the story of the brain we get from the physical sciences. But the story we get from neuroscience involves neuronal mechanisms underlying various forms of observable behaviour, and this story doesn't seem to have a place for subjective feelings like pain. So in order to fit pain into that story, the proponent of the Brain First approach ends up redefining pain, say, in behavioural terms: to be in pain is to behave in a pained way, i.e. to instigate avoidance behaviour as the result of bodily damage.

Suddenly the problem of consciousness looks easier. It's still incredibly difficult of course, but it seems like we have some kind of grip on how physical brain processes could produce pain behaviour; and if that's all that's involved in pain then it looks like we can account for pain without adding to the story told by the physical sciences. The problem is that in redefining pain it also looks like we've changed the subject: we're no longer talking about the inner feeling of pain; we're now talking about pain behaviour.

The Russellian insights outlined in the last section enable us to turn the problem of consciousness on its head, adopting what we might call a 'Consciousness First' approach to the problem:

Consciousness-First Approach – Our first-person understanding of consciousness is not only a datum in its own right, but a datum to be used for understanding the nature of matter. We build up our metaphysical picture of the brain not only by examining what we know 'from the outside' about the brain's causal structure, but also by reflecting on what we know 'from the inside' about its intrinsic character. Attention to experience provides direct insight into the intrinsic character of the brain, and given that we have no other method for directly accessing the intrinsic character of matter, this insight provides our best clue as to the nature of matter in general.

It is hard to overstate what a radical shift this is in our conception of reality and our epistemological access to it. Current wisdom tells us that we know a lot about the physical world, but face a challenge seeing how consciousness fits into it. Russellian wisdom tells us we know a lot about consciousness, but face a challenge seeing how matter fits around it. If and when Russell's insight seeps into mainstream thinking, it will revolutionise our understanding of what natural science is, and of its role in finding out what the world is like.

References

Alter, T. & Nagasawa, N. (Eds.) 2015. *Consciousness and the Physical World*, Oxford University Press.

Armstrong, D. 1997. *A World of States of Affairs*, Cambridge: Cambridge University.

Bird, A. 2007. *Nature's Metaphysics: Laws and Properties*, Oxford: Oxford University Press.

Blackburn, S. 1990. 'Filling in space,' *Analysis* 50: 62–65.

Campbell, K. 1976. *Metaphysics: An Introduction*, Encino, CA: Dickenson.

Chalmers, D. J. 1996. *The Conscious Mind: Towards a Fundamental Theory*, Oxford University Press.

- Chalmers, D. J. 2009. 'The Two-Dimensional Argument Against Materialism,' in B. McLaughlin (Ed.) *Oxford Handbook of the Philosophy of Mind*, Oxford University Press, 313–39.
- Eddington, A. 1928. *The Nature of the Physical World*, Cambridge University Press.
- Ellis, B. 2001, *Scientific Essentialism*, Cambridge: Cambridge University Press.
- Ellis, B. 2002, *The Philosophy of Nature: A Guide to the New Essentialism*, Montreal: McGill-Queen's University Press.
- Heil, J. 2003. *From an Ontological Point of View*, Oxford: Clarendon Press.
- Jackson, F. 1982. 'Epiphenomenal Qualia,' *Philosophical Quarterly* 32.
- Jackson, F., 1986. 'What Mary Didn't Know,' *Journal of Philosophy* 83.
- Kirk, Robert. 1974. "Sentience and Behaviour", *Mind*, vol. 83, pp. 43–60.
- Lowe, E. J. 2006. *The Four-Category Ontology: A Metaphysical Foundation for Natural Science*, Oxford: Oxford University Press.
- Molnar, G. 2003, *Powers: A Study in Metaphysics*, Oxford: Oxford University Press.
- Mumford, S. 2004. *Laws in Nature*, London: Routledge.
- Papineau, D. 2001. 'The Rise of Physicalism' in B. Loewer (Ed.) *Physicalism and its Discontents*, Oxford University Press.
- Robinson, H. 1982. *Matter and Sense*, Cambridge: Cambridge University Press.
- Russell, B. 1927. *The Analysis of Matter*, Kegan Paul.