Is Realism about Consciousness Compatible with a Scientifically Respectable World View?

A response to Keith Frankish’s ‘Illusionism as a Theory of Consciousness’

Illusionism is the view that the apparent reality of phenomenal consciousness is a powerful illusion; illusionists deny that the anything really has phenomenal consciousness. This is a position which deserves exploration, and Keith Frankish’s recent defence of it is rigorous and compelling. It also brings together, perhaps for the first time, the varied literature on this topic.

Many philosophers will protest that illusionism is obviously false, or that the assertion of it is somehow how self-defeating. I am inclined to think both of these things. However, my aim in this piece is not to argue that illusionism cannot be true, but rather to undermine the motivation for it. Frankish believes that realism about phenomenal consciousness is at odds with scientifically respectable metaphysics. I hope to show that this is not the case, or at least that Frankish has given us no grounds for thinking that it is.

Is Illusionism coherent?

Phenomenal consciousness is typically defined as follows:

Something is phenomenally conscious if and only if there is something that it’s like to be it.

In his paper Frankish articulates a deflationary understanding of ‘what it’s like’ talk, and hence for the sake of clarity we can add that phenomenal consciousness involves experience of qualities: the redness of red experiences, the bitter quality involved in the experience of biting into a lemon, the qualitative character of heat sensations. Indeed it is precisely the qualitative character of consciousness that Frankish finds problematic. In what follows I will use the word ‘consciousness’ to refer to phenomenal consciousness.

The illusionist accepts that human beings represent states of consciousness, and indeed that we typically believe that states of consciousness exist. But she also holds that all such representations are non-veridical, and the corresponding beliefs false. Just like beliefs about the Loch Ness Monster, beliefs about consciousness fail to correspond to anything real.

Beliefs are a kind of propositional thought. Is it coherent to accept the reality of thought whilst denying the reality of consciousness? That depends on whether or not there is a constitutive relationship between thought and consciousness. Frankish assumes throughout the paper that we can account for thoughts, such as beliefs and other mental representations, without the postulation of consciousness. In this he follows the dominant view in analytic philosophy that there is no essential connection between thought and consciousness. This view was largely unquestioned in the twentieth century. However, there is now a growing movement in analytic philosophy defending the thesis that thoughts, and indeed mental representations in general, are identical with (or directly

1 There were a few opponents this consensus, e.g. John Searle (1984), Galen Strawson (1994) and Charles Siewart (1998).
constituted of) forms of phenomenal consciousness. Uriah Kriegal has dubbed this movement ‘the Phenomenal Intentionality Research Program.’

Clearly if the convictions of the Phenomenal Intentionality Research Program turn out to be correct, then illusionism involves a straightforward contradiction: you can’t assert the existence of thought but deny the existence of consciousness if thought just is a (highly evolved) form of consciousness. I believe there is strong reason to accept that thought is a form of consciousness, and hence strong reason to think that illusionism is indeed incoherent. However, I will not have space to defend this here. I merely note that illusionism depends on an assumption which Frankish has not defended in this paper: that mental representation is not dependent on phenomenal consciousness. In what follows I will grant him this assumption for the sake of argument.

**Is Consciousness Anomalous?**

Frankish’s principal motivation for illusionism is his conviction that consciousness is an ‘anomalous’ phenomenon, at odds with our scientific picture of reality. It might be useful to distinguish two ways in which a certain thesis might be in tension with our scientific knowledge of the world:

*Being Anomalous by Inconsistency* – A thesis is anomalous by inconsistency iff it is inconsistent with what we know (or have good reason to believe) about reality through empirical methods. An example of this is the creationist thesis that the world is between 6,000 and 10,000 years old.

*Being Anomalous by Irreducibility* – A thesis is anomalous by irreducibility iff it requires ontological commitments over and above those required by what we know (or have good reason to believe) about reality through empirical methods. A plausible example of this is the thesis that there is a non-interventionist God.³

At various points Frankish seems to suggest that consciousness is anomalous is both of these senses. In the next section I will argue that the reality of consciousness is not anomalous by inconsistency. I will then go on to argue that, although there may be grounds for thinking that consciousness may be anomalous by irreducibility, Frankish has given us no reason to think that this fact is in any way problematic for the consciousness realist.

**Is Phenomenal Consciousness Anomalous by Inconsistency?**

Why might one think that the reality of consciousness is inconsistent with what we know about reality scientifically? Frankish supports the standard arguments of Chalmers, Jackson and others to the conclusion that states of consciousness are non-physical.⁴ This in itself does not entail an inconsistency with what we know empirically; the postulation of non-physical properties may merely add to what we know about the world empirically, without contradicting it. However, inconsistency may arise when we try to find a causal role for such non-physical consciousness in the material world.

Paul Churchland, among others, has argued that giving a causal role to non-physical properties violates principles of energy conservation:

---

² Kriegel 2013.
³ This thesis is anomalous by irreducibility only if design arguments for the existence of God fail.
⁴ Chalmers 1996, 2009, Jackson 1982, 1984. I critique these forms of the argument and suggest an alternative in Goff 2011 and Goff MS.
...as has been known for more than fifty years...forms of Dualism...fly in the face of basic physics itself...since any position that includes non-physical elements in the causal dynamics of the brain must violate both the law that energy is neither created nor destroyed, and the law that the total momentum in any closed system is always conserved. In short, you simply can’t get a change in any aspect of the physical brain (for that would causally require both energy changes and momentum changes) save by a compensatory change in some other physical aspect of the brain, which will thereby lay claim to being the cause at issue. There is simply no room in a physical system for ghosts of any kind to intervene in some fashion to change its dynamical behavior. Any physical system is ‘dynamically closed’ under the laws of Physics. (Indeed, it was this very difficulty, over a century ago, that initially motivated the desperate invention of Epiphenomenalism in the first place.)

This argument is far too hasty. As David Papineau has pointed out fundamental mental forces need not constitute a counterexample to the principle of the conservation of energy, so long as those forces act conservatively; so long as they ‘operate in such a way as to “pay back” all the energy they “borrow” and vice-versa.’

...the conservation of energy in itself does not tell which basic forces operate in the physical universe. Are gravity and impact the only basic forces? What about electro-magnetism? Nuclear forces? And so on. Clearly the conservation of energy as such leaves it open exactly which basic forces exist. It requires only that, whatever they are, they operate deterministically and conservatively.

We must distinguish this kind of argument, based on conservation of energy, from more common and more promising arguments based on the alleged causal closure of the physical. Frankish’s concerns about the causal role of phenomenal consciousness are of the latter kind:

Non-physical properties can have no effects in a world that is closed under causation, as ours appears to be, and the mind sciences show no independent need to refer to exotic physical processes, such as quantum-mechanical ones. The threat of epiphenomenalism hangs over such theories.

This familiar worry is perhaps most associated with Jaegwan Kim. If we grant the premise that consciousness is non-physical, then we can turn this worry into an argument that non-epiphenomenal consciousness is anomalous by contradiction:

1. The thesis of causal closure (i.e. the thesis that every physical event has a sufficient physical cause) is empirically well supported.
2. Consciousness is non-physical
3. If consciousness exists, and it is not epiphenomenal (i.e. it has a causal impact on the physical world), then there would be physical events (i.e. the events caused by consciousness) which have a non-physical rather than a physical cause, in violation of causal closure.

5 Churchland 2014.
6 Papineau 2002: 252.
7 P. 13
9 The truth of this premise relies on an implicit assumption that the effects of consciousness are not systematically over-determined. This premise is usually explicit in discussions of this issue, but I leave it implicit here in order to keep things simple.
Therefore, either consciousness is epiphenomenal or its existence is inconsistent with an empirically well supported thesis (i.e. causal closure).

The thesis that the physical world is causally closed is often stated, but very rarely defended. Following this trend Frankish does not defend, nor reference any defence of, the principle, beyond saying that it ‘appears to be true’. Peter Rauschenberger has recently given a powerful critique of the scant empirical defence of causal closure which has appeared in the literature.10

Moreover, as Frankish points out, even if causal closure is true, there is a way of reconciling the non-physicality of consciousness with the causal closure of the physical world: Russellian monism.11 Russellian monism results from a recent rediscovery of the approach to the mind-body problem defended (independently) by Russell and Eddington in the 1920s.12 The starting point for the view is Russell’s observation that the physical sciences, such as physics, chemistry and neuroscience, characterise physical properties dispositionally, i.e. solely in terms of what those properties do. In neuroscience, for example, a given brain state is characterised in terms of (i) its causal role in the brain, (ii) its chemical constituents. Those chemical constituents are then characterised in terms of (i) their causal role, and (ii) their physical constituents. And the basic physical constituents are characterised entirely dispositionally; mass, for instance, is characterised in terms of gravitational attraction and the disposition to resist acceleration.

Bearing this in mind, consider a case in which a certain brain state is strongly correlated with a certain conscious state. Suppose, for the sake of having an example, that we discover a strong correlation between hunger and ventromedial hypothalamus stimulation (VHS). Physical science tells us nothing about what VHS is beyond what it does (and what its constituents do). Now there are powerful arguments to the conclusion that there must be more to the nature of a property than its causal role, and if those arguments are successful then there must be more to VHS than what physical science tells us about it.13 But even if those arguments don’t work, it is surely coherent to suppose that there is more to VHS than its causal role, that VHS has a categorical nature which underlies its dispositional characteristics. If Frankish thinks that the rather orthodox philosophical view that dispositional properties are grounded in categorical properties is incoherent or implausible, then he owes us an argument to justify that conclusion.

The Russelilian monist has a proposal for what that the categorical nature of VHS is: the feeling of hunger. According to Russellian monism, physical science describes VHS ‘from the outside’, i.e. it tells us what it does; whilst in introspection we know VHS ‘from the inside’ as the conscious experience of hunger. VHS is a single state with a qualitative intrinsic nature which underlies the dispositional characteristics which are the focus of empirical science.

Now if pain is a non-physical state, and pain is the intrinsic nature of VHS, does it follow that VHS is a non-physical state? In fact, in the context of Russellian monism we need to note certain ambiguities in the term ‘physical property’. If Russellian monism is true, then VHS is referred to by the predicates of neuroscience, and in that sense VHS is ‘physical’; but it has a categorical nature which is not revealed by physical science, and in that sense it is ‘non-physical’.14

---

10 Rauschenberger MS. Papineau 2001 is perhaps the best defence of causal closure, although even here the empirical support is explored in very little detail.
11 For a recent collection of essays on Russellian monism see Alter and Nagasawa 2015.
12 Russell 1927, Eddington 1928.
14 Stoljar 2001 outlines a detailed account of these two definitions of physical.
It is in the latter sense that the standard ‘anti-physicalist arguments’, such as the zombie conceivability argument and the knowledge argument, try to demonstrate that consciousness is ‘non-physical’. But consciousness only needs to be ‘physical’ in the former sense in order to secure its causal role. If the conscious feeling of hunger is the categorical nature of ventromedial hypothalamus stimulation, then it is in virtue of being so that it is ensured a genuine causal role in the material world.

Frankish objects to Russellian monism on the grounds that ‘it involves huge profligacy with phenomenal properties and preserves the potency of consciousness only at the cost of making all physical causation phenomenal.’ However, this objection assumes that Russellian monism entails panpsychism, which is not the case. Although there are panpsychist versions of Russellian monism, such as the view of Strawson which Frankish references, there are also non-panpsychist versions. It is perfectly possible to be a Russellian monist whilst only believing in those forms of consciousness we have a pre-theoretical commitment to.

There is therefore no reason to suppose that there is any tension between full-blooded realism about non-epiphenomenal qualitative phenomenal consciousness and our empirical knowledge of the world; there are no grounds for thinking that non-epiphenomenal consciousness is anomalous by contradiction. Even if we have reason to believe that the physical world is causally closed, and this has not been adequately defended, there is way of reconciling causal closure with a commitment to conscious states with are both causally efficacious and non-physical (‘non-physical’ in the sense of having a nature not fully captured by physical science, and this is all that is demonstrated by the standard ‘anti-physicalist’ arguments).

Is it a Problem if Phenomenal Consciousness is Anomalous by Irreducibility?

Russellian monism allows us to accommodate full-blooded qualitative consciousness into the physical world without contradicting anything we know about the world empirically. But it does involve adding to what we know about the world empirically (I am here taking ‘empirical’ knowledge to be knowledge which results from third-person observation and experiment). Both the Russellian monist and the illusionist accept that states of phenomenal consciousness are not accessible to third person scientific methods, and are not explicable in terms of the dispositional properties that feature in physical science. If this is correct, then realism about consciousness takes us beyond the metaphysical commitments of third-person science.

Frankish assumes throughout the paper that it is problematic to go beyond the metaphysical commitments of third-person science. But why should we think this? It goes without saying that we should not believe in things that we have no reason to believe in. But the realist would claim that our apparent introspective awareness of our own consciousness does give us reason to believe in the reality of consciousness. If we have reason to believe in consciousness, and if there is good reason to believe (as Frankish thinks there is) that consciousness cannot be explained in terms of the postulations of third person science, it follows that we have reason to believe in something over and above the postulations of third-person science.

Of course, if there were a tension between the reality of consciousness and the reality of consciousness and the facts of natural science, then we would have to do some weighing of

---

15 P. 13.

16 For example, Pereboom 2011 defends a disjunction of non-panpsychist Russellian monism and illusionism. I favour a panpsychist version of Russellian monism (Goff MS), but those impressed by Frankish’s worries need not follow me in this.
epistemic reasons. However, we demonstrated in the last section that there is no such tension. Once we have dispelled the worry that consciousness is anomalous by contradiction, the onus is on the illusionist to demonstrate that our apparent introspective awareness of consciousness gives us no reason to believe in the reality of consciousness. Frankish accepts that it is epistemically permissible to believe in the external world on the basis of our senses. Why is it not epistemically permissible to believe in phenomenal consciousness on the basis of introspection?

Frankish points to empirical evidence that first-person judgements are not reliable. But it is a huge leap to go from the fact that our putative introspective access to phenomenal consciousness is not perfect, to the claim that it has no epistemic force at all. If there is no empirical case against the reality of consciousness, all the realist needs to hold is that our apparent introspective awareness of phenomenal consciousness gives us reasonable grounds to believe that phenomenal consciousness exists. Frankish is obliged to tell us why this belief is unreasonable.

There are of course epistemic disadvantages pertaining to our epistemic access to consciousness. Facts about phenomenal consciousness cannot be inter-subjectively checked in the way that facts about the external world can. This is unfortunate, and introduces all sorts of methodological difficulties to the science of consciousness. However, I can’t see an argument from the fact that our apparent access to phenomenal consciousness is non-ideal to the conclusion that it ought not to be trusted at all. Third-person access to the external world is non-ideal either; for example, there seems to be no way of purging it of sceptical doubt (Arguably our access to consciousness does not suffer from this drawback, but perhaps it would be question begging to assert that in this context). The epistemic situation of evolved creatures is significantly inferior to the epistemic situation of angels; we’ve got to take what we can get.

The most interesting arguments Frankish raises against the plausibility of consciousness realism focus on the peculiar epistemic relationship that seems to obtain between the mind and its conscious states. I will spend the rest of this section focusing on these arguments in some detail.

Frankish claims that ‘if realists are to maintain that phenomenal consciousness is a datum, then they must say that we have a special kind of epistemic access to it, which excludes any possibility of error’. I am not clear why he thinks that consciousness must be known with certainty in order to be a datum; the realist might merely claim that introspection provides good reason to believe in the reality of consciousness, just as the senses provide good reason to believe in the external world. Still, it is true that many realists, especially those who take consciousness to be irreducible, do believe that conscious subjects stand in a special, non-causal relationship of ‘acquaintance’ to their conscious states; and that in virtue of the relationship of acquaintance the existence and nature of conscious states is known with something close to certainty.

Frankish raises a few worries about the relationship of acquaintance:

1. **First, acquaintance can have no psychological significance.** In order to talk or think about our phenomenal properties we need to form mental representations of them, and since representational processes are potentially fallible, the certainty conferred by acquaintance could never be communicated.

---

17 P. 18.
18 Chalmers 2003, Goff 2015, MS.
19 P. 18.
In response to this worry we can note that if there is such thing as acquaintance, then it is itself a kind of representational process: in being acquainted with pain I represent that property to myself. And no consciousness realist that I know of holds that all representations of consciousness are certain or infallible; David Chalmers for example holds that only a narrow class of mental concepts – those he calls ‘direct phenomenal concepts’ – have this special status. Thus, I can’t see why any realist would worry that the certainty involved in acquaintance-based representations of consciousness is lost when the content of the concept is incorporated into other representational processes. Again Frankish seems to be assuming that the reality of phenomenal consciousness is a datum only if it is known with certainty, and I fail to see what reason we have to think that. My experience of the table in front of me does not guarantee the table’s existence, but it is nonetheless reasonable to believe in the table on the basis of it.

Second, acquaintance theory assumes that reactions and associations a sensory episode evokes do not affect its feel, since we are not directly acquainted with them or their effects. Yet there is reason to think that our reactions and associations do shape our sense of what our experiences are like.20

It would be good to hear more detail about how the evidence Frankish references counts against an acquaintance model. For the moment I am not clear why the acquaintance theorist cannot accept that certain association and tendencies to react do have a causal impact on phenomenal states, phenomenal states which are subsequently known in acquaintance-based representations. And given that the acquaintance theorist holds that only a limited subset of representations of consciousness involve certainty, she is quite entitled to hold that reactions and associations cause a certain amount of misrepresentation among our representations of consciousness. Perhaps there are some empirical worries about acquaintance, but I think we need to hear more detail.

Acquaintance theory also comes with heavy metaphysical baggage. It is hard to see how physical properties could directly reveal themselves to us... Moreover, it may require an anti-physicalist view of the experiencing subject too. If subjects are complex physical organisms, how can they become directly acquainted with consciousness?21

This seems to be an Ockhamist objection to realism about direct acquaintance relations and subjects which stand in such relations. But if we have reason to believe in consciousness, and we have reason to believe that consciousness involves subjects bearing direct acquaintance relations, then we have reason to believe in subjects bearing direct acquaintance relations. If there was a clash between acquaintance and our empirical knowledge of person and brain, then we might have grounds for rejecting acquaintance. But cognitive science and neuroscience are limited to tracking the dispositional features of brains. It could be that the brain also has a categorical nature, and that that categorical nature involves a subject bearing direct acquaintance relations. If we have reason to believe in consciousness, and this is what consciousness involves, then this is what we ought to believe (in the absence of a powerful counter-reason).

To summarise my argument so far:

- Frankish has given us no reason to believe that phenomenal consciousness is anomalous by inconsistency. Even if the physical world is causally closed, and this has not been adequately defended, Russellian monism provides a way of reconciling causal closure with full-blooded realism about phenomenal consciousness.

20 P. 18.
21 P. 19.
If the conceivability argument or the knowledge argument are sound, then realism about phenomenal consciousness requires metaphysical commitments over and above those of third-person observation and experiment. However, if we have reason to believe in phenomenal consciousness, then we have epistemically legitimate grounds for making those additional commitments. And Frankish has given us no good reason to doubt that our apparent introspective access to phenomenal consciousness gives us good grounds for believing in the reality of phenomenal consciousness (just as our apparent sensory access to the external world gives us good grounds for believing in the external world).

**Radical Naturalism**

Why is Frankish so convinced that we have no good reason to believe in consciousness? I suspect that the fundamental motivation for Frankish, and for the other illusionists he discusses, is a deep commitment to a radical form of methodological naturalism, which we can define as follows:

*Radical naturalism* – Third-person observation and experiment, and only third-person observation and experiment, should be our guide in finding out the nature of reality.

The reality of consciousness does not seem to be accessible to third-person observation and experiment; I can’t see your pain no matter how much I poke around in your head. And thus if we accept radical naturalism, illusionism becomes extremely plausible. The question we now need to ask is: What reason do we have to be radical naturalists?

I believe that the attraction to radical naturalism arises from an emotional response to the success of science. It cannot be denied that in the last five hundred years or so the project of natural science has gone extremely well. From the movement of the planets, to the evolution of life, to the fundamental constituents of matter, natural science seems to be an unstoppable juggernaut of explanation. For the radical naturalist, what this shows is that we’ve finally found something that works, something we can put our ontological faith in. For thousands of years before the scientific revolution philosophers struggled to find out what reality is like and got nowhere. Since the scientific revolution natural science has enjoyed success after success after success. From this perspective, philosophers who look somewhere other than third-person empirical science to try to work out what reality is like are ‘old school’, dragging us back to the dark ages. They are to be equated with folk who believe in magic, or deny climate change, or think that the world was created in six days.

However, there is a deep irony here. A key moment in the scientific revolution was Galileo’s declaration that mathematic should be the language of natural science. But Galileo felt able to make this declaration only once he had stripped the material world of *sensory qualities*. For Galileo, colours, tastes, smells and odours aren’t really in *matter*, but in the *experiencing soul* of the perceiver. By stripping matter of its qualities, leaving only size, shape, location and motion, Galileo gave us, for the first time in history, a material universe which could be exhaustively described in the austere language of mathematics.

Thus, natural science begins with Galileo limiting its scope of enquiry, by putting the sensory qualities in the soul and entreating ‘natural philosophers’ to focus only on what can be captured mathematically. This limited project has gone extremely well, allowing the construction of mathematical models which yield extremely accurate predictions of the behaviour of matter. This in

---

22 In contrast to Descartes, Galileo conceived of the soul in Aristotelian terms, as the principle of animation in the body (corpo sensitivo). See Ben-Yami 2015 chapter 3 for more discussion of this.

Turn has enabled us to manipulate the natural world in extraordinary way, and thus to produce technology undreamt of by previous generations. This incredible success has created in contemporary philosophers a great desire to place all of their ontological faith in natural science.

However, it is clear upon calm reflection that the fact that things go really well we ignore the sensory qualities gives us absolutely no reason to think that sensory qualities don’t exist. It would be nice if we could apply the methods of third-person science to the qualities of experience, but their private nature in incompatible with public observation, and (as Galileo realised, and Frankish concurs) their qualitative nature resists (exhaustive) mathematical characterisation. This doesn’t mean they don’t exist; it just means that the human epistemic situation is far from ideal. We have at least as much reason to believe in the qualities of experience – on the basis of introspection – as we do to believe in the external world – on the basis of perception. And if we want a complete theory of the world, then we must try to construct one which incorporates everything we have reason to believe in, not just those things which are easier to deal with.

Natural science has done an extremely good job of describing the causal structure of matter. The job of the metaphysician is to build on this, by developing a theory which unifies the findings of natural science with other things we have reason to believe in. What other things do we have reason to believe in? At the very least the experiential qualities which Galileo set outside of the scope of natural science.

The project of metaphysics is currently hampered by an understandable but irrational attraction to scientism, which results from the visceral impact of technology on philosophical inclinations. So much so that the general public has almost no involvement with the metaphysics which goes on in philosophy departments, nor even awareness that it is going on. At some point this irrational attraction to scientism will fade, and society as a whole will return to the noble project of trying to work out what reality is like; this time armed with rich information about the causal structure of reality, information which was not available to our ancestors.

Many bemoan the fact that metaphysics doesn’t seem to have got anywhere. I’m inclined to think it hasn’t really begun.


Rauschenberger, P. (MS) The Dogma of Causal Closure.


