

Katalin Balog

Illusionism's Discontent

Abstract: Frankish positions his view, illusionism (a.k.a. eliminativist physicalism), in opposition to what he calls radical realism (dualism and neutral monism) and conservative realism (a.k.a. non-eliminativist physicalism). Against radical realism, he upholds physicalism. But he goes along with key premises of the gap arguments for radical realism, namely, 1) that epistemic/explanatory gaps exist between the physical and the phenomenal, and 2) that every truth should be perspicuously explicable from the fundamental truth about the world; and he concludes that because physicalism is true, there could not be phenomenal truths. I think he is wrong to accept 2; and even if he was right to accept it, the more plausible response would be not to deny the existence of qualia but to deny physicalism. In either case, denying the existence of qualia is the wrong answer. I present a physicalist realist alternative that refutes premise 2 of the gap argument; I also make a general case against the scientism that accompanies Frankish's metaphysics.

Keith Frankish has spelt out an interesting case for illusionism by positioning it as the least unpalatable of three rather defective positions on consciousness. The other two are what he calls *radical realism* (which covers dualism and neutral monism) and *conservative realism*, a.k.a. non-eliminativist physicalism. His view, which can be described as *conservative eliminativism*, denies the radical realist's contention that the world, and especially those portions of it that have minds around, consists of more than just physical stuff; but also takes issue with conservative realism (non-eliminativist physicalism) despite sharing its commitment to a purely physical ontology. Against conservative realism, he argues that it is not possible to be a realist about consciousness and still remain a physicalist; and he thinks our reasons for physicalism override our reasons to believe in qualia. My

Correspondence:

Email: kbalog@andromeda.rutgers.edu

own perspective is just the reverse. I hold that our reasons to believe in qualia are stronger than our reasons for physicalism — but I also think that qualia realism is compatible with physicalism.

My remarks will be mostly critical; but I want to register that I have found Frankish's views not only very engaging, but also extraordinarily helpful in orienting myself in the philosophical landscape. In the first section I will look at Frankish's general arguments for illusionism and what a conservative realist (i.e. non-eliminativist physicalist) can say in defence of realism. In the second part I will look at the plausibility of illusionism in its own right, and in the third highlight some specific problems with his account of the nature of the 'illusion'.

1. Physicalism vs. scientism

The most compelling consideration Frankish presents for illusionism is related to a well-known family of arguments, let's call them *gap arguments*, that aim to refute physicalism by appeal to various (epistemic, conceptual, and explanatory) gaps between physical and phenomenal descriptions of the world which, according to these arguments, provide *a priori* reason to reject physicalism. Frankish puts his concern with conservative (physicalist) realism in a very similar vein:

The central problem, of course, is that phenomenal properties seem too weird to yield to physical explanation. They resist functional analysis and float free of whatever physical mechanisms are posited to explain them. (Frankish, this issue, p. 25)

Here is how this leads to illusionism:

In general, apparent anomalousness is evidence for illusion. If a property resists explanation in physical terms or is detectable only from a certain perspective, then the simplest explanation is that it is illusory. In this light, considerations usually cited in support of a radical approach to consciousness, such as the existence of an explanatory gap, the conceivability of zombies, and the perspectival nature of phenomenal knowledge, afford equal or greater support for illusionism. (pp. 27–8)

After aligning himself with the gap arguments for illusionist conclusions, Frankish continues to hammer away at realism. He thinks our introspective judgments about conscious experience should be discounted because

...through external inspection of our brain states, they appear to be non-veridical; the properties represented do not show up from other perspectives. (p. 28)

These last remarks expose an important aspect of Frankish's illusionism. He combines the crucial premises of the gap argument — in this case, that all truths should be perspicuously explicable from the fundamental truths, and the observation that there are no physical explanations of qualia — with physicalism to argue that qualia don't exist. This combination of views supports and feeds on his *scientism*: that the best way to study everything is through science. Though scientism goes beyond the metaphysical position Frankish explicitly argues for, it is a plausible extension of it and is certainly in the background of the views of other notable illusionists Rey (1995) and Dennett (1991).

I think scientism is, for all the wonderful progress science has made, a theoretical mistake; but one with adverse practical consequences. I find the denial of qualia utterly implausible, and scientism a misguided way to approach one's life. There is a concern, expressed in Western philosophy most forcefully by Kierkegaard, namely that our experience of life matters in ineffable ways that no objective understanding of the world can capture. Wittgenstein, in a well-known letter to Ludwig von Ficker, the publisher of the *Tractatus*, claimed that 'the whole point of the book is to show that what is important lies in what cannot be expressed' in a scientific language (Wittgenstein, 1979, p. 94).

Suppose there was a super-intelligent organism — in a twist on Frank Jackson's (1982) knowledge argument — that lacked any feeling or experience, a creature of pure thought. A purely scientific account of humans — though very far from reality — is perhaps not an impossibility. So such a being could know everything about humans in biological, neuroscientific, and information processing terms — even though she lacked the introspective understanding normal humans have of their subjective reality. Such a creature would arguably know nothing of value, meaning, and human significance.

This is, of course, a crude opposition. We hardly ever relate to anything purely objectively or purely subjectively. But as Kierkegaard pointed out, modern life supports a tendency toward objectivity, while, in fact, one needs to become subjective, fully immersed in one's consciousness, to properly understand oneself. He said:

...this is the wonder of life, that each man who is mindful of himself knows what no science knows, since he knows who himself is...
(Kierkegaard, 1980, pp. 78–9)

So I reject scientism in any of its forms. I believe that if the existence of qualia was incompatible with physicalism it would be reason to

reject physicalism (rather than embrace it and scientism, as Frankish does). But I do not think that the existence of qualia and subjectivity is incompatible with physicalism. While I would rather give up physicalism before I'd embrace scientism and give up belief in first-person authority about the mind, I do not think this is a choice forced on physicalism. One can have physicalism without the scientism and illusionism about qualia. It is tenable that purely physical minds conceive of their own contents, and especially their (real) qualitative states in a way that is inaccessible to and isolated from the point of view of science. Of course, given physicalism, no facts exist that do not have an exhaustive third-person account, but this doesn't exclude first-person takes on the world and the mind that cannot be — perspicuously — explained from the scientific perspective.

This is exactly the perspective of what has become known as the 'phenomenal concept strategy' (Stoljar, 2005). The strategy is based on an idea first articulated by Brian Loar (1990; 1997) that the epistemic, conceptual, and explanatory gaps between phenomenal and physical descriptions can be explained by appeal to the nature of phenomenal *concepts*, thereby obviating both the illusionist and the anti-physicalist response to the gaps. Phenomenal concepts, on this proposal, involve unique cognitive mechanisms, but none that could not be fully physically implemented.

The key idea of the phenomenal concept strategy is to give an account of how phenomenal concepts can refer to conscious states *directly* and yet in a *substantive* manner, even while supposing that they refer to physical (plausibly, neural) states in the brain, via entirely physical mechanisms. On this view, both qualia and the phenomenal concepts we apply to them are physical; but phenomenal concepts involve unique cognitive mechanisms that set them apart — in fact, isolate them — conceptually from scientific concepts. Loar's core idea is that when a person is having a particular experience she can deploy a concept that refers directly to the experience and that in some way involves in its mode of presentation the very experience it refers to, and that this account of phenomenal concepts is entirely neutral with regard to the metaphysical status of conscious states; that is, entirely neutral on the question of whether qualia are physical or irreducibly mental. It also explains why physicalism about qualia *seems* to be puzzling.

One way to understand this idea, the one I favour, is to hold that phenomenal concepts are partly constituted by tokens of the phenomenal experiences they refer to (Balog, 2012a,b; Block, 2006;

Chalmers, 2003; and Papineau, 2002). On this view, a token phenomenal experience is *part* of the token concept referring to it, and the experience — at least partly — determines that the concept refers to the experience it contains a token of. Of course, ‘part’ does not mean ‘spatial part’ but rather that it is *metaphysically* impossible to token the concept without tokening an instance of its referent.

This account of phenomenal concepts is not intended to apply to all concepts that refer to phenomenal states but only to what we might call ‘direct phenomenal concepts’. Of course most of our reference to phenomenal states and qualia do not contain the phenomenal states themselves. Clearly, a person can token a concept that refers to pain without her literally experiencing pain — these can be called ‘indirect phenomenal concepts’ — as when she replies to her dentist’s question with ‘I am not in pain’ or when one sees another person stub her toe and thinks ‘that hurts’. But for the purposes of discussion it is appropriate to focus on direct phenomenal concepts since these are the ones that generate the puzzlement over qualia.

If the above account is right, phenomenal concepts have very special *realization states*: the neural states realizing these concepts are instances of the very same neural states types the concepts refer to. What is so special about phenomenal concepts, on this account, is not only that their *realization states* are instances of their *referents*, but that this very fact is crucially involved in determining their meaning. In other words, not only are these concepts realized by instances of their referents, but they refer to what they do at least in part *in virtue* of this fact. This is, of course, very different from any other concept. Most concepts are not realized by tokens of their referents at all; but even those — like the concept ATOM — that are, mean what they do completely independently of this fact about realization. This also means that the cognitive mechanisms involved in phenomenal concepts guarantee that we will be puzzled by how qualia fits in with the brain, whether or not physicalism is true.

Let me briefly explain how the constitutional theory of phenomenal concepts accounts for the explanatory gap. Recall that the problem of the explanatory gap is that no amount of knowledge about the physical facts (brain functioning and so on) is able to explain why a particular brain state/process has a particular feel, e.g. feels giddy. This contrasts, for example, with the way the fact that water is composed of H₂O molecules together with physical and chemical laws explains why water is potable, transparent, and so on. The explanation of why H₂O behaves in watery ways (together with the fact that water is

composed of H₂O molecules) straightforwardly explains the behaviour of water. Since we can't explain why a brain state feels giddy in neurophysiological terms, we can't close the physical–phenomenal explanatory gap.

The constitutional account explains the gap by appealing to the *substantial* and *direct* grasp phenomenal concepts afford of their referent. When I focus on the phenomenal state, I have a 'substantial' grasp of its nature. I grasp what it is like to be in that phenomenal state — in terms of *what it's like to be in that same state*. This is what the constitutional account captures. And because this grasp is at the same time *direct*, that is, independent of any causal or functional information (unlike in the case of WATER), information about the functioning of the brain simply won't explain *what it's like to be* in that state.

What exactly is this *substantial* insight into the nature of phenomenal states? If phenomenal concepts are partly constituted by phenomenal states, our knowledge of the presence of these states (when we apply these direct phenomenal concepts) is not mediated by something distinct from these states. Rather the state itself serves as its own mode of presentation. When I focus on the phenomenal quality of an experience — not on what it represents but on its qualitative character — my representation contains that very experience. Thinking about it and simply having the experience will then share something very substantial, very spectacular: namely the phenomenal character of the experience. Being aware of our phenomenal states — being *acquainted* with them (Russell, 1910) — is the special, intimate epistemic relation we have to our phenomenal experience through the *shared phenomenality* of experience and thought. Shared phenomenality produces the sense that one has a direct insight into the nature of the experience. And it seems puzzling, to say the least, how this nature could be physical. But it is important to notice that this kind of direct insight (via shared phenomenality of thought and experience) into the nature of conscious experience does not reveal anything about the metaphysical nature of phenomenality. It is not the same sense of 'insight into the nature of X' as a scientific analysis of a brain state would provide. The one involves *having* the state, the other, analysing it into its components, which are very different activities.

So the constitutional account of phenomenal concepts offers a solution to the mind–body problem that steers clear of both radical realism and illusionism. As a matter of fact, it tackles head on the main reason Frankish cites for illusionism, that is, the existence of epistemic/conceptual/explanatory gaps between the physical and the

phenomenal; and concludes that it is not a good reason to give up either physicalism or realism. Though Frankish mentions this approach, he dismisses it quickly, without much discussion.

In contrast, I think we should not take qualia lightly, and should look very seriously at views that could ground a physicalist realism about qualia. Illusionism should only be considered after all other avenues have been exhausted. Just as illusionism about the external world is hard to take seriously even though in a certain sense it fits our data well, illusionism about qualia should not be invoked lightly. Belief in the existence of qualia is just as foundational for our worldview as — and some would say even more so than — belief in the existence of the external world. It takes much more to make it a plausible position than simply showing that it is — at least *prima facie* — coherent and fits some other, initially plausible, principles.

2. The plausibility of illusionism

Frankish makes an attempt to neutralize the inherent implausibility of illusionism by explaining our stubborn (and supposedly erroneous) sense that we are phenomenally conscious. According to illusionism, when we are introspectively aware of our sensory states our awareness is partial and distorted, leading us to *misrepresent* the states as having phenomenal properties. So though nothing in this world, as a matter of fact, instantiates phenomenal properties, it still *appears* to us that our experiences do.

Frankish then tries to coax us to see his claim that introspection misrepresents as plausible. But the analogies he supplies fail to convince. He refers to Dennett's analogy with computer graphics.

The icons, pointers, files, and locations displayed on a computer screen correspond in only an abstract, metaphorical way to structures within the machine, but by manipulating them in intuitive ways we can control the machine effectively, without any deeper understanding of its workings. The items that populate our introspective world have a similar status, Dennett suggests. They are metaphorical representations of real neural events, which facilitate certain kinds of mental self-manipulation but yield no deep insight into the processes involved. (Frankish, this issue, p. 16)

As Frankish himself points out, this analogy is quite imperfect. Introspection is not like looking at a computer screen, and computer icons are not misrepresentations either. No one who uses a computer really

thinks that computer files are located on the screen or that they look like their icons.

Frankish also cites Rey (1995, pp. 137–9), who explains the illusion of qualia as being similar to other illusions where stabilities in our reactions to the world induce us to project corresponding properties onto the world (e.g. our stable personal concerns and reactions to others lead us to posit stable, persisting selves as their objects). Again, while this mechanism is contentious even as an explanation of our concept of self, it doesn't seem to be similar at all to how our introspective phenomenal concepts work. When we form an introspective phenomenal concept of a pain sensation in the act of attending to it, we are not conceptually engaging, much less projecting stable personal concerns and reactions; we simply mentally note the pain. We can be aware of qualia via simple direction of attention. Everybody can do this and there is nothing tendentious about it. It is just a bed-rock feature of what it is to be a human being.

Despite Frankish's examples and explanations, I find illusionism extraordinarily implausible simply because it flies in the face of one of the most fundamental ways the world presents itself to us: the awareness of our own mind. Illusionism perhaps sounds plausible, or at least conceivable, from the third-person, scientific perspective we can take on mental representation. From this point of view, it is possible to argue that organisms have no introspective way of checking the accuracy of their introspective representations, and so they cannot rule out the possibility that these representations are non-veridical.

It is clearly the case that science and objective philosophizing might dislodge deeply held common sense views. Obvious examples are the nature of physical objects, and, more controversially, the nature of the self and free will. But the case of qualia is not like that. Arguably, *pace* Frankish, there are no scientific or philosophical discoveries that force us to give up belief in qualia; and there is no demonstrable conceptual incoherence in our introspective concepts of qualia. So the question comes down to the epistemic authority accorded to introspective awareness vs. scientific theorizing.

It seems that Frankish has a negative view of qualia and their role in our life. He, in the strange expression he uses, finds qualia potentially *embarrassing*. He thinks that illusionism can eliminate the embarrassment and clear away the obstacles from taking qualia seriously. As he puts it:

But if phenomenal properties are intentional objects, a sort of mental fiction, then we need no longer be embarrassed by them. We can

acknowledge how magical and unearthly they are *and* how powerfully they affect us, as intentional objects. In this sense, illusionists may claim to take consciousness more seriously than realists do. (p. 29)

But there are also signs that Frankish is not completely at ease with illusionism.

3. An illusion of illusion?

Some aspects of Frankish's presentation of illusionism strike me as covert attempts to smuggle qualia in through the back door. He seems to me to appeal — illicitly — to qualitative properties in explicating and motivating his own denial of them. First I will talk about problems regarding reference to non-existent qualia, then I will make some remarks about Frankish's treatment of what it is like to have an experience.

3.1. Phenomenal concepts

The heart of illusionism is the view that introspection misrepresents sensory experience as having certain qualitative properties nothing in fact has. But given how vivid our grasp of these allegedly uninstantiated properties are, one is owed an explanation how, and through what mechanism, we can latch onto something that doesn't exist in such a revealing way. The story, of course, cannot run along the same lines as the story for our concept 'unicorn' does; our phenomenal concepts are simple and direct in a way that precludes construction from other, *bona fide* referring concepts. As Levine (2001, pp. 146–7) has observed, there appears to be a problem accounting for the infinitely rich ways in which these concepts apparently refer to an infinitely rich field of phenomenal properties. It is very challenging to explain what it means to represent phenomenality directly — if there is no such thing.

Frankish's answer doesn't come close to meeting the challenge:

A better option may be to adopt some form of functional-role semantics for phenomenal concepts, on which their content is fixed by their role in mental processing, including their connections to other concepts, to non-conceptual sensory and introspective representations (their own content determined causally or functionally), and to associations, behavioural dispositions, and so on. (p. 36)

This is, unfortunately, little more than hand-waving about how reference to non-existent (or non-instantiated) properties with direct

modes of presentation can be established. And in fact there is reason to be suspicious that such an account could ever be found. The problem can be stated as a dilemma. Either introspective concepts refer to real properties so introspection results in meaningful even though erroneous representations, or they don't really refer to any property. In the first case, one just wonders what miracle could ensure that people refer directly to all those wonderful qualitative and subjective properties even though nothing in the world instantiates them? And in the second, all our introspective qualia representations would simply be meaningless, mental junk, so to speak. So the account either requires a miracle, or collapses into meaninglessness.

I suspect that there is a tacit appeal to qualia in illusionism which makes it initially plausible. Because in reality we are all acquainted with qualia, we don't get worried about the idea that introspective representations can refer to them. But when we realize what the account says, namely that nothing has qualia, it should really strike us as utterly miraculous that, *if the account was true*, we could refer to them.

3.2. *The 'what it is like' of experience*

While illusionism denies the existence of qualia, Frankish seems to want to allow that there is something it is like to think about experience, and even talks about 'introspective subjectivity'. One might wonder: where is the illusion then? It would be pointless to deny that experience has qualitative, subjective properties only to allow introspective representation of experience to have them. That would still be a realist position. So the illusionist's 'what it is like' must be construed otherwise than as 'having qualitative features'. Here is how Frankish explains the distinction:

Illusionists can say that one's experiences are like something if one is aware of them in a functional sense, courtesy of introspective representational mechanisms. Indeed, this is a plausible reading of the phrase; experiences are like something for a creature, just as external objects are like something for it, if it mentally represents them to itself. Illusionists agree that experiences are like something in this sense, though they add that the representations are non-veridical, misrepresenting experiences as having phenomenal properties (what-it's-likeness in the first sense). (p. 23)

Even assuming that it is supposed to be a constitutive account of what-it's-like-ness, this is not very helpful. To create a new sense of 'what

it is like', it not only has to be different from 'having qualia properties', but it also has to be discernably different from concepts of mere function and representation — otherwise invoking the expression 'what it is like' is just a funny way to dress up 'function' and 'representation' talk. It is a redefinition of the concept 'what it is like', rather than a new understanding of it. It introduces no new insight about experience.

Nevertheless, Frankish seems to think that his new concept of 'what it is like' really does speak to our ordinary notion of what it is like. He uses the account to dispel misconceptions about zombies:

[The illusion] depends on a complex array of introspectable sensory states, which trigger a host of cognitive, motivational, and affective reactions. If we knew everything about these states, their effects, and our introspective access to them, then, illusionists say, we could not clearly imagine a creature possessing them without having an inner life like ours. (p. 23)

Assuming that this is not merely a claim about imagination, but about conceivability, this indicates that, according to Frankish, suitable claims about representation and function conceptually necessitate claims about what it is like to be a creature entertaining those representations. This would indicate that he indeed provided a functional-representational notion of what-it's-like-ness — not merely a pseudo-what-it's-like-ness concept. But this flies in the face not only of what most philosophers believe about phenomenal concepts (that they do not have conceptually sufficient conditions in functional/representational terms), but also of the main reason Frankish presented for illusionism: the gap arguments. If zombies are unimaginable, indeed inconceivable, then the case for radical realism vanishes and conservative realism becomes a viable option.

So I think Frankish overplays his hand with his claim that he can account for the what-it's-like-ness of experience. It might be that, even for an illusionist, the allure of qualia is too strong to resist. But trying to have his cake and eat it will in philosophy, as in the kitchen, get you into trouble when members of your family arrive.

References

- Balog, K. (2012a) Acquaintance and the mind-body problem, in Hill C. & Gozzano, S. (eds.) *The Mental, the Physical*, pp. 16–43, Cambridge: Cambridge University Press.
- Balog, K. (2012b) In defense of the phenomenal concept strategy, *Philosophy and Phenomenological Research*, **84** (1), pp. 1–23.

- Block, N. (2006) Max Black's objection to mind-body identity, in Zimmerman, D. (ed.) *Oxford Studies in Metaphysics*, II, pp. 3–78, Oxford: Oxford University Press.
- Chalmers, D. (2003) The content and epistemology of phenomenal belief, in Smith, Q. & Jovic, A. (eds.) *Consciousness: New Philosophical Perspectives*, pp. 220–273, Oxford: Oxford University Press.
- Dennett, D.C. (1991) *Consciousness Explained*, New York: Little, Brown.
- Frankish, K. (this issue) Illusionism as a theory of consciousness, *Journal of Consciousness Studies*, **23** (11–12).
- Jackson, F. (1982) Epiphenomenal qualia, *Philosophical Quarterly*, **32**, pp. 127–136.
- Kierkegaard, S. (1980) *The Concept of Anxiety*, Thomte, R. (trans. & ed.), Princeton, NJ: Princeton University Press.
- Levine, J. (2001) *Purple Haze: The Puzzle of Consciousness*, Oxford: Oxford University Press.
- Loar, B. (1990) Phenomenal states, *Philosophical Perspectives*, **4**, pp. 81–108.
- Loar, B. (1997) Phenomenal states: Second version, in Block N., Flanagan, O. & Guzeldere, G. (eds.) *The Nature of Consciousness: Philosophical Debates*, pp. 597–616, Cambridge, MA: MIT Press.
- Papineau, D. (2002) *Thinking About Consciousness*, New York: Oxford University Press.
- Rey, G. (1995) Towards a projectivist account of conscious experience, in Metzinger, T. (ed.) *Conscious Experience*, pp. 123–142, Exeter: Imprint Academic.
- Russell, B. (1910) Knowledge by acquaintance and knowledge by description, *Proceedings of the Aristotelian Society*, **11**, pp. 108–128. Reprinted in Russell, B. (1963) *Mysticism and Logic*, pp. 152–167, London: Allen and Unwin.
- Stoljar, D. (2005) Physicalism and phenomenal concepts, *Mind and Language*, **20** (2), pp. 296–302.
- Wittgenstein, L. (1979) Letters to Ludwig Ficker, in Luckhardt, C.G. (ed.) *Wittgenstein: Sources and Perspectives*, Gilette, B. (trans.), Ithaca, NY: Cornell University Press.