Illusionism and anti-functionalism about phenomenal consciousness

Derk Pereboom, Cornell University


Penultimate draft

Abstract. The role of a functionalist account of phenomenal properties in Keith Frankish’s illusionist position results in two issues for his view. The first concerns the ontological status of illusions of phenomenality. Illusionists are committed to their existence, and these illusions would appear to have phenomenal features. Frankish argues that functionalism about (quasi-) phenomenal properties yields a response, but I contend that it doesn’t, and that instead the illusionist’s basic account of phenomenal properties must be reapplied to the illusions themselves. The second concern is that phenomenal properties would seem to be intrinsic properties of experience, but functionalism has them consist solely in relations. The non-functionalist option can recapture the sense that these properties are intrinsic. It can also preserve the intuition that they are causal powers in a robust sense, and thereby, perhaps surprisingly, provide a stronger response to the knowledge and conceivability arguments.

1. Introduction

I generally agree with Keith Frankish’s very fine exposition and defense of the illusionist option about phenomenal consciousness (for perhaps the earliest version see Dennett 1988; cf. Churchland 1981), and with his account of its theoretical advantages over radical and conservative realism. As Frankish notes, there are strong exclusionary
pressures against dualist radical realism. Such a radical realism may deny the causal closure of the physical, but not without cost. And I agree that conservative realism, in particular the phenomenal concept strategy version, is subject to the kinds of instability concerns that Frankish sets out, although the issues are complex and undoubtedly more remains to be said. There are two issues on which I want to comment, and both concern the functionalist interpretation of the illusionist option. That is, both concern whether the strongest sort of illusionism opts for the view that phenomenal properties or their quasi-phenomenal correlates, to use Frankish’s term, are best cast as consisting exclusively of causal relations to sensory inputs, behavioral outputs, and other mental states. I generally favor a non-functionalist account of the mental for the reason that it accords best with a non-Humean account of causal powers generally, including mental causal powers (Pereboom 1991, 2011).

Functionalism has two roles in Frankish’s account that I will challenge. The first concerns the ontological status of illusions of phenomenality themselves. Illusionists are committed to their existence, and these illusions would appear to have phenomenal features. The second is that phenomenal properties would seem to be intrinsic properties of experience, but functionalism about these properties would have them consist solely in relations, that is, extrinsic features. The non-functionalist option can preserve the sense that the reality accessed is in fact an intrinsic feature of experience, and this, I argue, has a number of advantages.

2. Do illusions have phenomenal properties?

So first, illusionists agree that the what-it-is-like features of sensory states, what I
call the qualitative natures of phenomenal properties, are illusory in that they don’t exist. But what of the illusion itself? Experiencing an illusion requires that the illusions themselves, which are kinds of introspective representations, will exist. As Frankish himself remarks, “Forming the theoretical belief that phenomenal properties are illusory does not change one’s introspective representations” (p. 7). However, these illusions themselves will have phenomenal properties, or so it would seem. Illusions of phenomenal properties would appear to differ in a key respect from other sorts of illusions, for instance from the illusions of psychokinesis that Frankish adduces. Illusion of psychokinesis will not have psychokinetic properties, while illusions of phenomenal properties would seem to have phenomenal properties. Accordingly, the objector may claim that there will be something it’s like to have an illusion of phenomenal greenness, and it’s the same as what it’s like have a sensation of green. If this is right, then the illusionist’s strategy won’t get us very far (see Pereboom 1994: 582-84, Alter 1995, and Chalmers 1996: 142 for versions of this objection).

In setting out his account of the concern, Frankish says: “The objector may reply that in order to create the illusion of a greenish experience, the introspective representation would have to apply a greenish mode of presentation, which would itself have an introspectible greenish feel.” In his response he contends that “illusionists will simply deny this, arguing that the content of introspective representations is determined by non-phenomenal, causal, or functional factors.” (p. 20). However, this seems insufficient. The objector may even agree that the content of the illusory experience is functional, but point out that the illusion nevertheless exists, and argue there is something it’s like to have it.

Frankish points out in a footnote (p. 20, n. 30) that I set out an answer to this
objection that reapplies the illusionist’s hypothesis to such modes of presentation. I think that this answer is essential to strength of the illusionist’s case. The proposed alternative, claiming functionalist content for the illusory representation, appears not to address the objection effectively.

Let me explain my response in some detail, and I’ll begin with a summary of my way of setting out illusionism. The core concern about physicalism expressed in the both Frank Jackson’s (1980) knowledge argument and David Chalmers’s (1996, 2002) conceivability argument begins with the claim that phenomenal states have phenomenal, what it-is-like type, properties, and concerning them it seems plausible that:

(i) Introspective modes of presentation represent a phenomenal property as having a specific qualitative nature, and the qualitative nature that an introspective mode of presentation represents a phenomenal property as having is not among the features that any physical mode of presentation would represent it as having.

Moreover, it is also seems plausible that:

(ii) The introspective mode of presentation accurately represents the qualitative nature of the phenomenal property. That is, the introspective mode of presentation represents the phenomenal property as having a specific qualitative nature, and the attribution of this nature to the phenomenal property is correct.

One way to characterize the qualitative natures that introspective modes of presentation represent phenomenal properties as having is by way of resemblance to modes of presentation (cf. Locke 1689/1975): Mary’s introspective representation of her phenomenal-red sensation presents that sensation in a characteristic way, and it is plausible that a qualitative nature that resembles this mode of presentation is accurately
attributed to the sensation’s phenomenal property. Alternatively specified, this qualitative nature is just as the introspective mode of presentation represents it to be.

Given these claims about what is least initially plausible, the proponent of the knowledge argument can explain its force in the following way. Mary, upon leaving her black-and-white room and seeing the ripe tomato, comes to have the belief:

(A) Seeing red has R,

where ‘R’ is the phenomenal concept that directly refers to phenomenal redness, that is, to phenomenal property R. We’ll assume that the qualitative nature of phenomenal redness is accurately represented introspectively by Mary’s introspective phenomenal mode of presentation. We’ll also assume that if physicalism is true, then every truth about the qualitative nature that an introspective mode of presentation accurately represents a phenomenal property as having will be derivable from a proposition detailing only features that physical modes of presentation represent the world as having (an assumption I don’t challenge). However, (A) is not derivable from this proposition. So not every truth about the qualitative nature that introspective modes of presentation accurately represent phenomenal properties as having is so derivable. Thus physicalism about phenomenal properties is false.

This version of the knowledge argument can be challenged by questioning one or both of the claims about what is intuitive. The illusionist takes issue with (ii), the claim about the accuracy of introspective phenomenal representation. Supposing the truth of (i), it is an open possibility that introspective representation is inaccurate in the sense that it represents phenomenal properties as having qualitative natures they in fact lack; that is, it is an open possibility that the qualitative inaccuracy hypothesis is true. Upon seeing the red
tomato, Mary introspectively represents the qualitative nature of phenomenal redness in the characteristic way, but it is an open possibility that she is representing phenomenal redness as having a qualitative nature that it actually does not have. Here, then, is the illusion.

As Frankish notes, the seriousness of this open possibility can be supported by an analogy with our visual representation of colors (Pereboom 1994, 2011). Our visual system represents colors as having certain specific qualitative natures, and it is an open possibility, widely regarded as actual, that colors actually lack these natures. It’s open that introspection of phenomenal properties involves representational systems relevantly similar to those involved in visual color perception, and this would explain how the qualitative inaccuracy hypothesis about the phenomenal could be true. It’s plausible that part of what explains qualitative inaccuracy about color representation is that it’s causal, and this allows for a discrepancy between the real nature of colors and how we represent their qualitative natures. It might well be that introspective representation of phenomenal color is also causal, and that this gives rise to a similar discrepancy. A terminological and perhaps substantive point: Frankish opts for the view that the what-it-is-like qualitative nature is what phenomenality is. So what remains given inaccuracy and the resulting illusionism is a quasi-phenomenal correlate or a quasi-phenomenal cause of the introspective representation. But, as Frankish notes, abandoning the view that physical objects have properties that resemble our visual representations of color doesn’t motivate that color correlate that remains is not really color but only ‘quasi-color.’ Rather, color turns out to differ from what we might have naturally assumed it to be. Analogously, the illusionist can argue that phenomenal properties are preserved given her view, but they
turn out to differ from what we ordinarily assume them to be. But even if this is right, Frankish’s terminology is valuable for clarity of exposition, and I will use it in what follows.

Supposing that this illusionist open possibility is in fact realized, how should we characterize what happens when Mary leaves the room and sees the red tomato? She now comes to have a belief of the form:

(A) Seeing red has R.

Consider first the proposal that the concept ‘R’ in this belief refers to a property with the qualitative nature accurately represented by that characteristic mode of presentation. Supposing our open possibility is actual, seeing red lacks this qualitative nature, and as a result belief (A) will be false. Then coming to believe (A) wouldn’t amount to Mary’s learning something about the qualitative nature of phenomenal redness, that is, to acquiring a new true belief about it. Alternatively, consider the proposal that “R’ refers to a quasi-phenomenal property, which is introspectively misrepresented and lacks such a qualitative nature, and is uncontroversially entirely physical. We can suppose that this belief is then true, but given this supposition Mary already had this belief when in her room, or was able to derive it from the true beliefs then had, she also does not acquire a new true belief. So either Mary’s belief (A) about seeing red wouldn’t be true or wouldn’t be new.

It is at this point that we encounter the objection that illusions of phenomenality themselves have phenomenal properties. Thus it might be objected that the account as explained up to this point merely shifts the problem for a physicalist explanation from accounting for phenomenal properties to accounting for the introspective phenomenal modes of presentation that characterize illusions of phenomenal properties. Given the
qualitative inaccuracy hypothesis, the object of Mary’s sensation indeed lacks the qualitative nature that would be accurately represented by her introspective representation, and thus she cannot learn that this object has such a qualitative nature. However, she nevertheless seems to learn something about how the illusion’s mode of presentation – let’s call it ‘MPR’ – represents what it does. And it’s at least initially plausible that Mary cannot derive the corresponding truth about how MPR represents from her microphysical base.

Here is my response to this objection (2011, pp. 27-28; cf. Pereboom 1994). If the qualitative inaccuracy hypothesis applies to introspective representations of (quasi-) phenomenal states, it makes sense to suggest that it also applies to introspective representations of phenomenal modes of presentation. It would then also be an open possibility that Mary introspectively represents MPR as having a qualitative nature it really lacks. Then, despite how MPR is introspectively represented, she would be able derive every truth about its real nature from her microphysical base even while in her black and white room.

When I’ve presented this response to various audiences, I almost always encounter the objection that it generates an unwelcome infinite regress of some kind. For example, one might suspect that given this response, when I represent an introspective representation, I do so by way of a mode of presentation, whereupon I would represent this mode of presentation by another introspective mode of presentation, and I would represent that further mode of presentation by a yet further introspective mode of presentation, \textit{ad infinitum}. Thus when I represent an introspective representation, I would in fact have an infinite series of introspective representations, and this is absurd.
The plausibility of my response to the objection that illusions of phenomenality are themselves phenomenal requires that when Mary introspectively represents her sensation of red, she also introspectively represents the introspective mode of presentation MPR of that sensation, while no actual infinite regress of introspective representations is generated. A key point is that a mode of presentation can function as the way a subject represents something without that subject also representing the mode of presentation itself. She might, in addition, represent this mode of presentation, but this would be a distinct representation that is not necessitated by her initial representation. If she did represent the mode of presentation, it could be by means of a higher-order introspective mode of presentation. However, it would not be necessitated that she also represents this higher order mode of presentation.

In addition, it’s credible that when someone introspectively represents a sensation of red by way of MPR, she will normally, although not necessarily, also represent MPR by a higher order mode of presentation, but only in unusual cases will she introspectively represent that higher order mode of presentation, or modes of presentation at yet higher orders. Moreover, as Nico Silins suggested to me (in conversation), it is implausible that beyond some fairly low level of iteration mental states would be introspectively represented by way of phenomenal modes of presentation. At some level of iteration, I am able only to form a belief, without any distinctive phenomenology, that I am representing a mental state. Such a belief, because it is not phenomenal, would fail to lend support to the knowledge argument.
Thus by reapplying the illusionist’s basic strategy to the illusion itself, a key objection to illusionism can be answered, and given this reply, no actual infinite regress is generated.

3. The advantages of retaining intrinsic quasi-phenomenal properties.

My second concern is that phenomenal properties would seem to be intrinsic properties of experience, but illusionism supplemented by functionalism about phenomenal or quasi-phenomenal properties would have them consist solely in relations to causes and effects. The non-functionalist option can respect the intuition that phenomenal properties are in fact intrinsic features of experience, and also honor the sense that they are non-Humean causal powers. In addition, it provides a further response to the knowledge and conceivability arguments as set out by Jackson and Chalmers. True, illusionism already yields a response to these anti-materialist arguments without invoking non-functionalism. But the additional response bolsters the case against these arguments, and my sense is that preserving intrinsicalness makes the overall view more plausible.

Let me begin with the response non-functionalism provides to the anti-materialist arguments. Chalmers’s (2002; cf. 1996) exposition of his conceivability argument allows us to see this advantage most clearly: Let ‘P’ be a statement that details the complete physical truth about the actual world; ‘T’ a ‘that’s all there is’ statement, specifying that P describes a minimal P-world; and ‘Q’ an arbitrarily selected actual phenomenal truth. Statement S is ideally conceivable when it is conceivable on ideal rational reflection, and prima facie conceivable when conceivable on initial reflection. S is primarily possible just in case it is true in some world considered as actual. S is secondarily possible just in case S is true in
some world considered as counterfactual. Correlatively, S is primarily conceivable just in case S can be conceived as true in some world considered as actual, or alternatively, since considering-as-actual is an a priori matter, S is primarily conceivable just in case the subject can’t rule out S a priori. Given these preliminaries, here is the argument:

(1) ‘PT and ~ Q’ is ideally primarily conceivable.

(2) If ‘PT and ~ Q’ is ideally primarily conceivable, then ‘PT and ~ Q’ is primarily possible.

(3) If ‘PT and ~ Q’ is primarily possible, then ‘PT and ~ Q’ is secondarily possible, or Russellian monism is true.

(4) If ‘PT and ~ Q’ is secondarily possible, materialism is false.

(5) Materialism is false or Russellian monism is true.

Russellian monism is any view that combines (1) *categorical ignorance*, the claim that physics leaves us ignorant of certain categorical bases of physical dispositional properties with (2) *consciousness-or experience-relevance*, the proposal that these categorical properties have a significant role in explaining consciousness or experience.

The Russellian Monist escape is key to the advantage I have in mind. Here is how Chalmers conceives this escape. While primary/a priori conceiving of P fixes all of the dispositional and relational properties designated by physical concepts, it does not determine categorical and intrinsic properties that underlie and explain them. So it’s open that one can primarily conceive ‘PT and ~ Q’ only because one is conceiving just dispositional and relational properties on the physical side, and it is an open possibility that if one were to replace ‘P’ with a more complete ‘P*’ that includes concepts that allow for direct representation of the currently unknown or incompletely understood intrinsic
properties in their categorical basis, ‘P*T and ~ Q’ would not be primarily conceivable. ‘Q’ would then be a priori derivable from ‘P*T,’ and P* would explain phenomenal consciousness (Chalmers 2002, 2003; Stoljar 2006; Pereboom 2011; Alter and Nagasawa 2012).

The following is a crucial issue for the success of the argument: Suppose we agree that:

(0) ‘PT and ~ Q’ is prima facie primarily conceivable.

Can we secure:

(1) ‘PT and ~ Q’ is ideally primarily conceivable?

Without (1), we can’t advance to primary possibility. But we’re not ideal conceivers, so there’s a problem. But according to Chalmers, what licenses the move from (0) to (1) is the thesis that from truths solely about structure only truths solely about structure can be derived – the from-structure-only-structure thesis (Chalmers 2003; Stoljar 2006; Pereboom 2011; Alter 2016). Structural properties are, on first pass, relational, i.e., extrinsic. (Dynamical properties, also invoked in this context, are alterations in structural properties over time, and they are also relational, and we can think of them as a kind of structural property.) On a first, rough interpretation of this thesis, it tells us that only truths solely about relational properties can be derived from truths solely about relational properties.

On the assumption that the physical is purely structural and thus relational, and that phenomenal properties are intrinsic properties of experience and thus not relational, the relevant truths about phenomenal properties won’t be derivable from physical truths, no matter how much we come to know about the physical. Even in the ideal case, the relevant phenomenal truths won’t be derivable from physical truths.
But this clearly won’t work, since much of our physical knowledge is in fact about intrinsic and thus non-relational properties. For example, a ball’s shape and rigidity are intrinsic properties, and physics does not leave us ignorant of them. At this point I’ve propose the following refinement (2011, 92-94). The categorical bases of which physics leaves us ignorant (on Chalmers’s Russellian Monist conception) are intrinsic properties of a fundamental kind – *absolutely intrinsic properties*.

P is an *absolutely intrinsic* property of X just in case P is an intrinsic property of X, and P is not fully grounded in extrinsic properties of parts of X.

By contrast

P is a *(merely) relatively intrinsic* property of X just in case P is an intrinsic property of X, and P is fully grounded in extrinsic properties of parts of X.

(For X to ground Y, X must necessitate X and Y must exist in virtue of Y.) The notion of an absolutely intrinsic property facilitates a more plausible and precise characterization of the from-structure-only-structure thesis. Consider an objection that Daniel Stoljar raises against this principle:

The simplest way to see that the from-structure-only-structure thesis is false is to note that one can derive the instantiation of an intrinsic property from a relational one just by shifting what thing you are talking about. For example, being a husband is a relational property of Jack Spratt, and being a wife is a relational property of his wife. But being married is an intrinsic property of the pair (or the sum) of Jack Spratt and his wife. To take a different example, it seems plausible to say that I have the property of having a hand intrinsically, but my having this property obviously
follows from a relation between my hand and the rest of my body, and that the truth
concerning this is a relational truth. (Stoljar 2006: 152)

Torin Alter (2009) agrees that Stoljar has a point: if objects $x$ and $y$ compose object $z$, then
it is possible to derive intrinsic properties of $z$ from relational properties of $x$ and $y$. But
Alter argues that this objection undermines the from-structure-only-structure thesis only if
nonstructural properties are identified with intrinsic properties, and in his view that
identification is mistaken, for “the property being married is purely structural/dynamic
despite being intrinsic to the Spratts. Any structural/dynamical duplicate of the actual
world contains a corresponding married pair.” Alter contends that such examples indicate
not that we should reject the from-structure-only-structure thesis, but rather that we
should not identify nonstructural properties with intrinsic properties.

How might being a married pair be a structural property of the Spratts? The
distinction between relatively and absolutely intrinsic properties supplies a diagnosis.
Although being a married pair is an intrinsic property of the Spratts, it is fully grounded in
Jack’s extrinsic property of being married to Jill and Jill’s extrinsic property of being married
to Jack. Consequently, being a married pair is a relatively intrinsic property and not an
absolutely intrinsic property of the Spratts. I propose, then, that all nonstructural
properties are absolutely intrinsic properties or have absolutely intrinsic components.
Stoljar’s counterexample would then fail against the from-structure-only-structure thesis,
and, more generally, this principle would be in the clear (Pereboom 2011, 110-14).

But notice that given that the from-structure-only-structure thesis is what facilitates
the transition from

$$(0) \text{ PT and } \sim Q \text{ is prima facie primarily conceivable}$$
a non-structural conception of the physical would block it. Chalmers allows Russellian Monist conceptions of the physical, which would function as an escape from the argument, but the examples he adduces are restricted to versions of pan- or micropsychism (e.g., Strawson 2003) or involve currently unconceived intrinsic properties that are similar enough to paradigmatic physical properties to count as physical themselves. Illusionists might well resist any such proposals as being too mysterious. But there is a non-mysterious proposal for absolutely intrinsic physical properties to which illusionism per se wouldn't be averse. It crucially features a non-Humean, realist view of causal powers.

Causal powers are often cast as dispositional properties. But dispositions can be characterized as mere tendencies, to be explicited solely in terms of subjunctive conditionals. For a model, take the tendency of a ball to roll when pushed to be a dispositional property. The ball has a disposition such that if it were located on a plane surface and pushed, it would roll. It’s often argued that dispositions conceived in this way require categorical bases to explain the dispositional tendencies. The ball’s shape and rigidity qualify as such categorical bases. Typically, those categorical bases would be intrinsic properties of the entity that has them. On the non-Humean view, causal powers are identified with the intrinsic categorical bases, and not with the mere tendencies.

Jonathan Jacobs (2011; cf. John Heil 2003) explains this non-Humean notion of a causal power as that of an intrinsic property, not itself explicable in terms of subjunctive conditionals, that serves as a truthmaker for such subjunctive conditionals. The ball’s shape and rigidity, both intrinsic properties, are truthmakers for why it is that if the ball were
located on a plane surface and pushed, it would roll. On Jacobs’s view, this grounding in such a truthmaker explicates how it is that the categorical basis is a causal power. One might call such a categorical basis, following Sydney Shoemaker (in conversation), an *intrinsic aptness* to produce certain effects.

In Chalmers’s conceivability argument we’re thinking of P as the complete microphysical truth. Let’s agree that the most fundamental forces that current physics invokes – gravitation, electromagnetism, and the strong and weak nuclear forces – are described by physics in exclusively dispositional/relational terms, and that those dispositional descriptions are cast in terms of subjunctive conditionals such as: if an electron were in the vicinity of another electron, it would repel that electron. On Jacobs’s account of causal powers, they are intrinsic properties that serve as truthmakers for such subjunctive conditionals, and we can call them intrinsic physical aptnesses to produce specific effects. We can now add that these truthmakers would consist at least in part of *absolutely* intrinsic properties. And with such non-mysterious non-structural properties in the base (P*), the from-structure-only-structure principle won’t license the move from prima facie to ideal conceivability of ‘P*T and ~ Q’.

Thus if the illusionist were to accept that phenomenal or quasi-phenomenal properties consist at least in part of absolutely intrinsic aptnesses, she’d have a stronger response to the conceivability argument (and to the knowledge argument as well, since it implicitly requires the claim that Mary, as ideal reasoner, can’t derive ‘Q’ from ‘PT’, and we, as non-ideal reasoners, would need the from-structure-only-structure thesis in order to ascertain this claim). But this is inconsistent with functionalism about such properties, for then they would consist solely in relations. The illusionist would also have a response to
the claim that phenomenal properties are intrinsic properties of experience, which she can now accept. Finally, it’s intuitive that phenomenal properties are causal powers. Phenomenal pain and phenomenal pleasure would seem to be powers to cause aversive and attractive responses. It’s arguably intuitive, at least for many, that phenomenal are causal powers on the non-Humean model. If the reality of phenomenal pain consisted in a functional property, it would consist solely in relations to causes and effects, and then it could not be a causal power on the non-Humean model. To the extent that it is intuitive that pain is such a causal power, the non-functional account of phenomenal or quasi-phenomenal properties has the advantage.

References


Frankish, Keith. (2016). “Illusionism as a theory of consciousness.” This journal.


