

THE ABOLITION OF PHENOMENA: A VOYAGE AMONG THE ZOMBIES

Katalin Balog
(published in *Klésis* n° 55, 2023)

Abstract

This paper is about « strong illusionism », the view that we are not conscious at all, that there is nothing it is like, in the usual sense of those words, to feel sad, or to smell lavender. I argue against this view on a priori grounds; I also rebut the main philosophical arguments marshalled in its defense. And finally, I argue that illusionism doesn't only pose a theoretical problem, it is practically problematic as well since it undermines the moral concerns of its adherents. Therefore, one should resist illusionism not only in academia but in the wider culture.

Illusionism about consciousness is on the march: the view that our beliefs about consciousness are illusory is spreading from the halls of academia to popular media¹. This view is an attack on the last remnants of the premodern concept of the mind. Not so long ago, most people in the West thought of the body as mortal flesh commanded by the soul. They thought of the soul as immortal, free in its action and exempt from the laws of nature. Creativity and intelligence were believed to be traits of the soul that no mere mechanism could replicate. Dreams and visions were considered important messengers about some aspects of reality. All this is quite intuitive, and probably deeply ingrained in the everyday manner in which we still think about ourselves. But over the course of the last couple of hundred years, and especially in the 20th century, it has become common knowledge that the mind and the brain are tied together in a systematic and exceptionless way, and as the principle of the causal completeness of physics became mainstream among philosophers and scientists², it also became common understanding that all of our behavior has a purely physical explanation. The recent rise of artificial intelligence challenges our belief in the uniqueness of human creativity. There is not much about the pre-modern conception that survived these changes, except the view that we are conscious, i.e., that there is something it is like to be us, and that we are directly aware of it. It is precisely this view that illusionism attacks.

It is useful to pause here for a minute and briefly recap the dialectical landscape in which illusionism is the latest arrival. Philosophy responded to these changes by – mostly – jettisoning the Cartesian view of the soul as a simple immaterial substance with all its remarkable powers and alternating among three views each of which pose problems for philosophical understanding : physicalism, i.e., the view that mental properties, including consciousness, occur in virtue of complex arrangements of physical properties; dualism, which claims that mental states are over and above the physical; and panpsychism, according to which the intrinsic – as opposed to causal/dispositional – nature of fundamental physical properties and entities is conscious. All of these views are realist about the mind and consciousness though either deflationist about the nature of mind (physicalism) or deflationist about its causal role

¹ In philosophy, see, e.g., Dennett 1991, 2017; Rey 1995 ; Perebrom 2011, 2019 ; Kammerer 2016, 2021 ; Frankish 2016, in neuroscience, see Graziano 2016, 2019. The view is also making its way into popular culture. In a recent interview, for example, Sam Altman, CEO of OpenAI professed sympathy for the view.

² According to the causal completeness thesis, all physical effects are fully determined by a purely physical prior history. For a detailed description of how this view has become mainstream and a discussion of its philosophical consequences, see Papineau 2001.

(dualism and, to some extent, panpsychism). Illusionism takes a different tack: it claims that consciousness itself, the last redoubt of the mind, is altogether an illusion.

There is some uncertainty about what exactly this view comes to. Chalmers (2018) helpfully distinguishes *weak* from *strong* illusionism. Some illusionists are *weak illusionists* since they do not deny the existence of consciousness per se, i.e., they do not deny that there is something it is like to have experiences; instead, they deny some of our intuitions about their nature, for example, that conscious experiences are intrinsic, non-physical, primitive, ineffable, or non-functional. They find those views illusory. The illusion at play here is basically dualism about consciousness: weak illusionists are physicalist consciousness realists who deny dualism about consciousness. I have no contention with this view. In the bulk of this paper, I will defend a view of this sort.

My subject is strong illusionism³, the view that we are not conscious at all, that there is nothing it is like, in the usual sense of those words, to feel sad, or to smell lavender. Henceforth, I will call this view illusionism for short. According to illusionists, we are, in a technical sense, zombies. Illusionism is the logical endpoint of the gradual dethronement of the soul. It rejects dualism and panpsychism because the causal closure principle has awkward consequences for both regarding mental causation, and it rejects physicalist consciousness realism because it accepts the anti-physicalist arguments that consciousness cannot be physical. Illusionism prides itself on moving away from the last vestiges of our inflated, premodern view of ourselves. Just as it turned out to be a baseless, fictitious idea that we are immaterial souls capable of influencing matter, science tells us – so the illusionist argues – that consciousness is fiction as well.

This is, of course, a shockingly implausible idea; we are all very aware from our own case of the existence of consciousness. There is nothing else, one might argue, that we can know with as much certainty. But as absurd as it appears on its face, this view is gaining momentum. I will argue first that illusionism is a priori false given our understanding of how phenomenal concepts work. Later I will also argue, for those unconvinced by the a priori arguments, that the philosophical motivations for it are misguided, that physicalism indeed can account for this last « special » feature of the mind: consciousness. Despite what illusionists argue, one can defend physicalist consciousness realism – i.e., a physicalist position which does not deny that there is something it is like to experience things – in the face of anti-physicalist arguments. Once that is clear, the motivation for illusionism evaporates. While there is a stand-off between physicalist consciousness realism and non-interactive property dualism – the main competitors to account for the mind’s place in nature – in that each can accommodate common sense and scientific evidence equally well, illusionism loses fair and square.

One might think this is a harmless – or even edifying – academic dispute over the fundamental fabric of the world. But I think illusionism is not harmless at all. If allowed to seep into the way people understand themselves, it is likely to have a negative effect on people’s moral outlook. At the end of the paper, I will make a case for this.

1. Is illusionism a priori false?

Phenomenal concepts

³ E.g., Frankish (2012, 2016).

The heart of illusionism is the view that our phenomenal concepts – that is, the concepts we apply to our conscious states in introspection – misrepresent. They claim that introspection represents conscious sensory experience⁴ as having certain qualitative properties – i.e., having a certain phenomenology, or ‘what it’s like’ character – even though nothing *in fact* has these properties. But here, illusionism stumbles on a notable feature of phenomenal concepts. Phenomenal concepts represent by *exemplification*; they work by applying attention to an experience. This seems to me to be an indisputable feature of our phenomenal concepts: anybody who has the capacity to form them knows this. Phenomenal concepts leave no distance, so to speak, between themselves and the experience. Yet the illusionist view requires precisely such a distance between the introspective phenomenal concept of a pain as it occurs, and the pain itself; it requires the possibility of one occurring without the other.

I suspect that what makes this view seem even a little bit plausible is a tendency on the part of illusionists to think about phenomenal concepts not in the first-person, but in the third-person, objective mode. Illusionists first point out that concepts in general can misrepresent, and then propose that introspective concepts are a special case of this. They do not have a theory of how this happens. Instead, illusionists talk about introspection in objective, third-person terms like representation, conceptual role, nonconceptual sensory processing, etc., which make it appear as if there is no special problem there. But this merely distracts from the fact that, were we not familiar with those introspective concepts in the first-person mode, we would not understand what they are even talking about.

It is the first-person point of view that establishes the meaning of phenomenal concepts; third-person uses of them are parasitic on the meaning established in the first-person mode. The meaning of a phenomenal concept cannot be fully spelled out in terms of its functional role (either theoretical or folk-theoretical), or any other third-person accessible fact⁵. Schwitzgebel (2016) proposes a pre-theoretical definition of consciousness simply by examples; however, our grasp of such a definition still comes from introspection. Though ‘consciousness’, ‘phenomenality’ or ‘what-it’s-like’-ness are terms of art in public language, their meaning can be easily explained to anyone since in the privacy of our minds we have already employed phenomenal concepts many times.

So the first-person mode of putting matters with regard to illusionism cannot be simply sidestepped. But, once formulated in this way, the absurdity of the view becomes apparent. Because competence with phenomenal concepts is part of the mental life of normal humans – even of those who are not particularly reflective – the illusionist view is nearly impossible to believe. Put in the first-person, the illusionist claim is:

I am aware of q but q doesn’t exist,
where q is a conscious phenomenal experience. Even if out of an abundance of charity we formulate the thesis as:

I am aware of what appears to be q but q doesn’t exist,
it is still clear that q must exist since the token of the expression ‘what appears to be q’ that is used in my judgment is *formed* through my introspective awareness of what appears to be q. Therefore, I know that an experience that appears to be q exists, which means – since appearance and reality come to the same thing when it comes to experience – that q exists.

⁴ From here on, I will simply use « experience » for the states in question.

⁵ Stoljar (2005) argues that there are analyticities involving phenomenal concepts and that this refutes the idea that phenomenal concepts are direct and unanalyzable. However, his examples merely show that there are analytically necessary conditions on something being a conscious state; not that there are analytically sufficient conditions.

In fact, my knowledge of the existence of *q* is just a special case of Descartes's Cogito. The idea is that when I reflect on my own thinking (by which Descartes means any conscious state), its reality is the one thing I cannot be deluded about, even if I am deluded about everything else. I can come to know that I have a particular thought or experience solely based on my introspective awareness of this thought or experience. Reflection on what is involved in introspective awareness reveals why this is so. For example, when I reflect on my thought that new intelligent species might come into being, there is no room for error as to whether I have a thought at all, or about its content, since my awareness is partly constituted by that very thought. I know for sure that I am thinking this thought, and a fortiori that I am not thinking about maple syrup. Similarly with experience. When I reflect on my feeling angry, I know for sure that I am not a zombie (that I do have feelings) and that I am feeling angry (and not, e.g., languid). This is a priori knowledge in a loose sense; in the sense that I don't need empirical evidence for it beyond having the experience itself, and no further empirical evidence – e.g., evidence coming from the neuro-sciences – could dislodge it. It is an example of what Descartes calls « clear and distinct » understanding.

Of course, one can deny this – in the way one can deny knowledge of mathematical truth as well – by calling into question our own clear and distinct ideas. Illusionists would have us believe that consciousness is a giant hoax, something like the hoax of Descartes' Evil Demon conjuring up the illusion of a world, only worse. Our knowledge of our own mind seems Demon-proof but it isn't, it is *in fact* – and not merely in possibility – nothing but illusions. But this is surely going too far. If such skepticism spread to our knowledge of what we are *thinking* at any given moment, thinking itself would break down. If I can't be sure of what I am thinking right now, the cogency and relevance of any further thoughts I might produce would come into question. The illusionist, of course, could claim that doubting introspection only applies in the case of consciousness. However, the illusionist would still have to explain how the illusion that the awareness of a conscious state contains that very conscious state comes about – and how it is relevantly different from the case of thought – which would require a lot more than gesturing at the possibility of a causal breakdown between representations and what they represent. Recent literature on the unreliability of introspection in general – whereas it raises reasonable doubts about issues such as, e.g., the knowledge of our motives, the quality of peripheral vision, the existence pre-reflective self-awareness, or imageless thought – simply does not apply to the case of awareness of presently attended experience.

The problem is not only that illusionism doesn't have an account of how the illusion of the containment of the experience within the introspective awareness comes about; they also have no explanation of how we could form concepts of non-existent properties. In other cases where a concept is said to have no referent (e.g., 'ghost', 'flogiston', etc.) we have an explanation of how we have come to erroneously refer to the non-existent entity or property, typically involving a story about how certain appearances are produced in ways that do not satisfy the concept. But this won't work in the case where the concept in question is the concept of an appearance, therefore leaving no room between appearance and reality. The story cannot run along the same lines as our concept « unicorn » does either; our phenomenal concepts are simple and direct in a way that precludes construction from other, bona fide referring concepts. Given how vivid and direct our grasp of these allegedly uninstantiated properties are in introspection, we are owed an explanation how, and through what mechanism we can latch onto them if they doesn't exist. As Levine (2001, p. 146-7) has observed, there appears to be a problem accounting for how these concepts apparently refer to an infinitely rich field of conscious properties. It is very challenging to explain what it means to represent consciousness directly – if there is no such thing.

Frankish (2016) attempts an answer :

« [an] 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 nonconceptual sensory and introspective representations (their own content determined causally or functionally), and to associations, behavioral dispositions, and so on. »

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. The problem can be stated as a dilemma. Either introspective concepts refer to real (but uninstantiated) properties so introspection results in meaningful even though erroneous representations; or they don't really refer to any property. In the first case, one wonders what miracle could ensure that people refer directly to all those different conscious properties even though nothing in the world instantiates them? A gesture toward functional role semantics doesn't begin to answer this question. No pain, no gain, so to speak ; there has to be more to the illusionist story. Even if, on the physicalist assumption, everything can be explained in physical-functional terms, this doesn't relieve the illusionist of her particular explanatory burden. And in the second case, where introspective concepts do not refer to anything at all, all our introspective representations of conscious properties would simply be meaningless mental junk which they definitely do not appear to be.

It is a tacit appeal to consciousness which makes illusionism initially plausible. Because in reality we are all acquainted with conscious states, we don't get worried about what it is that illusionists are actually talking about. But when we realize what the account says, namely that nothing has consciousness, we should all be mystified how, *if the account was true*, we could still understand what they are talking about.

A point of clarification. Of course, if physicalism is true, « I am introspecting q but q doesn't exist » can be, at least in principle, translated into the third-person language of function and process; and stated like that it is not a priori false. But the fact that the statement is not a priori false when formulated in third-person language is not enough to save the illusionist. Because although it is not a priori false in every formulation, if it is a priori false in *some* formulation then it is false. 'Water= \neq H₂O' is not a priori false. But once we realize that it expresses the same proposition as 'Water= \neq water' which *is* a priori false, we know that it is false.

2. There are no good arguments for illusionism

Some notable eliminativists – which is what illusionists used to call themselves –, have argued that the existence of consciousness is not compatible with science (Dennett 1988 ; Rey 1995). But these arguments have serious problems. Recent illusionists rely more on philosophical arguments to the effect that physicalism is incompatible with the existence of consciousness. There are two kinds of argument of this sort, but, as I will show, neither one of them is able to establish the failure of physicalism. I will now present and rebut these arguments in turn.

The gap arguments

The most compelling consideration illusionists present for their views is related to a well-known family of arguments, let's call them gap arguments, that aim to refute physicalist phenomenal realism by appeal to various (epistemic, conceptual, and explanatory) gaps between physical and phenomenal

descriptions of the world. The idea is that these gaps – together with metaphysical principles that spell out the metaphysical consequences of these gaps – provide a priori reason to rule out consciousness in a purely physical world. Frankish (2016), for example, puts his concern with physicalist phenomenal realism like this : « 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. » (p. 13)

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. » (p. 16)

A simple version of Chalmers' 'zombie argument' – one of the anti-physicalist arguments⁶ – goes like this :

- 1) P&~Q is conceivable.⁷
- 2) If P&~Q is conceivable then P&~Q is metaphysically possible.
- 3) If Q is true and P&~Q is metaphysically possible then physicalism is false.
- 4) Q is true.

4) Physicalism is false.

The illusionist turns this argument into a modus tollens by upholding premises 1-3 but denying the consequence (i.e., affirming physicalism). The conclusion is that there are no phenomenal truths, i.e., that phenomenal consciousness does not exist.

- 1) P&~Q is conceivable.
- 2) If P&~Q is conceivable then P&~Q is metaphysically possible.
- 3) If Q is true and P&~Q is metaphysically possible then physicalism is false.
- 4) Physicalism is true

5) Q is false.

My response to these arguments is based on what Stoljar (2005) calls the *phenomenal concept strategy*. The phenomenal concept strategy was first articulated by Brian Loar (1990/1997) who argued that the epistemic, conceptual, and explanatory gaps between phenomenal and physical descriptions can be explained by an appeal to the nature of phenomenal *concepts*, and not by real *metaphysical* gaps between the physical and the phenomenal. The strategy calls into question premise 2 linking conceivability with possibility which both anti-physicalists and illusionists are relying on. Phenomenal concepts, on this

⁶ Similar arguments include, among others, arguments based on conceivability considerations by Kripke (1972), Nagel (1974), Bealer (1994), Chalmers (1996, and 2009), as well as the Knowledge Argument of Jackson (1982), versions of the Property dualism Argument in Robinson (1993), White (2007), and Nida-Rümelin (2007), and the Explanatory Gap Argument in Levine (2001) and (2007).

⁷ P is the complete fundamental physical description of the world, including the fundamental physical laws, and Q is a positive phenomenal truth, e.g., that someone is having a visual experience with a particular phenomenal character at a particular time.

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 consciousness and the phenomenal concepts we apply to them, are physical; but phenomenal concepts involve unique cognitive mechanisms that set these concepts apart – in fact, isolate them – from scientific concepts.

How could this be true? Let's think of our point in the previous section about the containment of experience in an act of introspective awareness. If we take this idea seriously, it follows that phenomenal concepts, that is, concepts of experience formed in introspection, such as a concept of feeling excited, are *constituted* by tokens of the conscious experiences they refer to, for example, by my present feeling of excitement (Block 2006; Chalmers 2003; and Papineau 2002). That is, a token conscious experience is *part* of the token concept referring to it, and it is in virtue of this relation that the concept refers to it. If that is so, the special concepts we can form of conscious experience in introspection are very different from the physical or functional concepts we use to describe the brain. This is why we are puzzled by how conscious states fit in with the brain, and this puzzlement occurs *irrespective* of whether physicalism is true. For all I have said about phenomenal concepts, it could be that both conscious experiences and the introspective concepts I am forming of them are simply brain states. But I won't be able to make this out from the way I think about them. When I attend to a feeling of excitement and form a phenomenal concept of it, I have a « substantial » grasp of its nature. I grasp what it is like to feel excited – in terms of *what it's like to feel excited*. And because this grasp is at the same time *direct*, that is, independent of any causal or functional information (unlike in the case of, say, concepts of brain states *as* brain states), information about the functioning of the brain simply won't explain why it is like *this* to feel excited.

There have been responses to the phenomenal concept strategy by anti-physicalists, most prominently by David Chalmers (2007)⁸. But as I will argue, Chalmers' response fails to grapple with the core idea of the phenomenal concept strategy on its own terms. Chalmers (2007) argues that phenomenal concepts are either not physicalistically explicable, or they cannot explain our epistemic situation regarding consciousness. For our purposes, I will understand 'epistemic situation' in terms of the conceptual, epistemic, and explanatory gaps between the physical and the phenomenal⁹. Chalmers sets up the following dilemma for the proponent of the phenomenal concept strategy :

If P&~C is conceivable, then C is not physically explicable.

If P&~C is not conceivable, then C cannot explain our epistemic situation.

where C stands for the claim that we possess phenomenal concepts with the relevant key feature (e.g., being constituted by an instance of the referent) posited by a physicalist account of phenomenal concepts.

The dilemma, as it is stated, however, is ambiguous. C, according to physicalism, can be conceptualized in different ways; it can be formulated in phenomenal language (C_{Phen}), or physical language (C_{Phys}).

⁸ Illusionists, such as Frankish (2016) also echo similar points.

⁹ I chose this understanding of 'epistemic situation' for simplicity and clarity though this doesn't track Chalmers' use in 2007 exactly; I believe, however, that the issues are the same either way.

Since conceivability, in all its varieties, is a conceptual matter, the evaluation of Chalmers' premises will depend on what conceptualization of C we have in mind.

Using this apparatus, we get the following four premises :

1_{Phen}) If $P \& \sim C_{Phen}$ is conceivable, then C_{Phen} is not physically explicable.

1_{Phys}) If $P \& \sim C_{Phys}$ is conceivable, then C_{Phys} is not physically explicable.

2_{Phen}) If $P \& \sim C_{Phen}$ is not conceivable, then C_{Phen} cannot explain our epistemic situation.

2_{Phys}) If $P \& \sim C_{Phys}$ is not conceivable, then C_{Phys} cannot explain our epistemic situation.

I am now going to evaluate these statements and see if Chalmers' reductio works, i.e., if he is able to show that something is wrong with the phenomenal concept strategy, since C cannot both be purely physical (or physically constituted) and explain our epistemic situation. In line with the phenomenal concept strategy, 1_{Phys} and 2_{Phen} will come out as vacuously true by virtue of having a false antecedent. Let me take them by turn. If we understand C_{Phys} as a truth given in fundamental physical language, 1_{Phys} comes out vacuously true, since arguably any true fundamental physical description of the world, e.g., C_{Phys} , is implied by the full fundamental physical description of the world P, so $P \& \sim C_{Phys}$ is not conceivable. As for 2_{Phen}, anybody who accepts the conceivability of zombies, as I do, will have to accept the conceivability of « phenomenal concept zombies » (i.e., creatures that are physically identical with us but have no phenomenal concepts) and so hold that $P \& \sim C_{Phen}$ is conceivable, which makes 2_{Phen} vacuously true. We must keep in mind all along that the phenomenal concept strategy is premised on the idea that zombies, and phenomenal concept zombies are conceivable under a phenomenal conceptualization but not under a physical conceptualization even though $P \& \sim C_{Phen}$ and $P \& \sim C_{Phys}$ express the same fact.

On the other hand, 1_{Phen} and 2_{Phys} have true antecedents, and they both have a consequent that appears at first sight damaging for physicalism. However, the right strategy is to embrace the apparently damaging conclusions; they turn out to be quite compatible with the success of the phenomenal concept strategy.

Let's take 2_{Phys} first. The physicalist can embrace the consequent of 2_{Phys}, namely that C_{Phys} cannot explain our epistemic situation. C must be cast in *phenomenal terms* for it to explain our epistemic situation with regard to the gaps. Our epistemic situation is characterized by a puzzlement over the gaps, i.e. over how things described in *phenomenal language* fit into a physical world described in *purely physical terms*. Both phenomenal and physical concepts figure essentially in the way we find these gaps puzzling. Consequently, only C_{Phen} – which conceives of phenomenal concepts in phenomenal terms – and not C_{Phys} , explains our puzzlement. It explains the existence of these gaps from the way phenomenal states are conceived phenomenally (by containment in states of introspective awareness). The trick is that this explanation provided in phenomenal terms about the constitution of phenomenal concepts is compatible with physicalism but cannot be stated in purely physical language. C_{Phys} doesn't have the conceptual resources to address how these gaps come about. Nothing about this shows that the phenomenal concept strategy is defective in any way.

Similarly, there is a perfectly good sense in which physicalists could affirm the consequent of 1_{Phen}, i.e., that C_{Phen} is not physically explicable, without any concessions to anti-physicalism. C_{Phen} is not physically explicable, in the very same way that the claim that some phenomenal state, like my feeling excited, occurs is not physically explicable. The sense of 'explicability' at play here is something like

perspicuous explicability, explicability without an explanatory gap. The phenomenal concept strategy provides a model of how lack of perspicuous explicability of some phenomenon in physical terms is compatible with the phenomenon being purely physical, or physically constituted. It doesn't promise to *close* the explanatory gap; it *explains* the existence of the gap. So it is not incongruous with physicalism that just as phenomenal truths are not physically explicable, truths about phenomenal concepts – when thought about in phenomenal terms – are not physically explicable either. What is conceded here – what Chalmers' argument succeeds at showing – is merely the existence of the epistemic gaps, *not* the existence of an ontological gap. Nothing further, and nothing more needs to be granted by the proponent of the phenomenal concept strategy.

The response to Chalmers' dilemma is that all horns of it, 1_{Phen} and 2_{Phys} included, are perfectly compatible with physicalism. Of course, Chalmers argues that if C_{Phen} is not physically explicable, then physicalism is not true. But this only follows if we assume something like premise 2 of the zombie argument which is precisely what the phenomenal concept strategy is calling in question. The phenomenal concept strategy provides a physicalist understanding of how the epistemic gaps invoked in 1_{Phen} and 2_{Phys} arise. While dualists explain the epistemic gaps by positing ontological gaps, relying on principles like premise 2 of the zombie argument, physicalist proponents of the phenomenal concept strategy deny premise 2 and its ilk, and explain the gaps – and all other puzzling features of consciousness – by reference to the cognitive features of introspection. This is the crux of the physicalist rebuttal of the anti-physicalist arguments.

Chalmers's rejoinder is that the explanatory scheme of the phenomenal concept strategy is circular. But there is nothing viciously circular about it. True, the explanation is only compatible with physicalism if we assume the falsity of premise 2. But if you give the physicalists that assumption, they will be able to explain why premise 2 nevertheless *appears* true, as well as explain all other puzzling phenomena surrounding consciousness, all perfectly compatible with physicalism.

In fact, Chalmers engages in the same kind of circular argumentation that he thinks physicalists are guilty of. He claims that there is no hope of the phenomenal concept strategy succeeding since the issue of whether the new gap – i.e., the one involving C_{Phen} and P – is compatible with physicalism can be raised with the same force as with respect to the original explanatory gap. That is, he defends the zombie argument from the phenomenal concept strategy by relying on premise 2, *the very same* principle that he is trying to defend. This argumentative strategy doesn't add anything new to the debate, just recycles the initial disagreement at the meta-level. Instead of marshalling new considerations against the phenomenal concept strategy, Chalmers' argument begs the question. It attempts to refute the strategy aimed against its central premise by simply assuming that premise to be true; the most that can be achieved that way is fight things to a draw.

This is a stalemate.¹⁰ Each side can unseat the other side's core assumption – if they are permitted to make one core assumption of their own. The anti-physicalist assumes premise 2 to be true, and explains the gaps by appeal to irreducibly mental, non-physical properties, arguing that no purely physical world could contain consciousness; the physicalist assumes premise 2 to be false and argues that phenomenal properties are purely physical or physically constituted, moreover, that there is a perfectly cogent explanation why it appears otherwise. Both can show that, once granted that one core assumption, their view is consistent and can rebut challenges from the other side. Neither side can, without begging the question against the opponent, show that the other's position is untenable. Where you end up depends on what you take as your starting point.

¹⁰ See Balog 2023 for an extended argument that there is a stalemate between the physicalist and anti-physicalist camp.

The physicalist can take this as license to hold phenomenal realism. There is no compelling reason to think that physicalism is incompatible with phenomenal realism (though if there were, dualism would be preferable to illusionism). It is interesting that while illusionists are willing to entertain the idea that our belief in consciousness is illusory, they don't consider the far more plausible idea that the principle anti-physicalists rely on, linking the epistemic gaps between the physical and the phenomenal to metaphysical gaps, is illusory. Illusionists rely on an earnest endorsement of the anti-physicalist core principles; they just draw the conclusion that consciousness doesn't exist, instead of deeming it to be non-physical.

The debunking argument

The other argument illusionists rely on is the debunking argument. It is based on the idea that we can explain our beliefs about why consciousness is special – e.g., that we are acquainted with it, that we can't explain it in physical terms, that we can conceive of zombies, that Mary learns something when she leaves the black-and-white room, etc. – in a way that is independent of consciousness. The argument, discussed though not endorsed in Chalmers 2018, goes like this:

1. There is a correct explanation of our beliefs about consciousness that is independent of consciousness.
2. If there is a correct explanation of our beliefs about consciousness that is independent of consciousness, those beliefs are not justified.

3. Our beliefs about consciousness are not justified.

Such explanations as the argument invokes of course do not yet exist; the argument is based on a promissory note. But regardless of the future state of cognitive science, I argue that this argument fails as well, and for reasons that are similar to those we discussed in connection with the zombie argument. I am not going to take issue with premise 2 of the debunking argument; it will suffice to focus my discussion on premise 1 which I will explain is false under any interpretation. The thing to keep in mind, again, is that explanation is a concept-dependent affair.

No one disputes that when the beliefs in question involve *intentional-phenomenal* terms, and so we think about them phenomenally, any explanation that doesn't appeal to consciousness or its cognates in phenomenal terms will miss the mark – this is just an instance of the explanatory gap between the phenomenal and the physical. But proponents of the debunking argument think that in order to not beg the question in favor of consciousness, these beliefs need to be stated in topic-neutral terms (Chalmers call these « quasi-phenomenal » terms), i.e., in terms that make no explicit appeal to consciousness or its cognates. They think that when stated in this way, we can explain them topic-neutrally, without any appeal to consciousness, and that, in their view, is enough for the purposes of the debunking argument.

Here is a – hopefully not too fanciful – example illustrating why this won't work. As I argue in (1999), there could be special concepts of brain states (I call them 'yogi concepts') that, described topic-neutrally, appear to be just like phenomenal concepts. They pick out brain states that they refer to directly, in the act of having of those states, as it were. The referent is, like in the phenomenal case, *present* in the thought, as much as the concept « presence » can be made sense of in a topic-neutral fashion. The difference is, the states they pick out are not phenomenal states, and yogis will admit that they have no idea about the precise nature of these states. These concepts work somewhat like blindsight concepts, except that they refer to brain states. Though we do not have such concepts, it

doesn't seem impossible for there to be thinkers with such concepts. And they seem to fit the outlines of the topic-neutral explanatory account Chalmers (2018) deems most promising (p. 34):

« We have introspective models deploying introspective concepts of our internal states that are largely independent of our physical concepts. These concepts are introspectively opaque, not revealing any of the underlying physical or computational mechanisms. Our introspective models attribute primitive mental relations to the qualities introspected. We seem to have immediate knowledge that we stand in these primitive mental relations to the qualities introspected. »

It seems like this profile fits both phenomenal concepts and yogi concepts; but thinking about experience gives rise to beliefs that we have some states with a special nature, whereas thinking about brain states via yogi concepts arguably does not. It follows that the shared features of phenomenal concepts and yogi concepts cannot explain our beliefs about consciousness, even the topic-neutral simulacra of these beliefs. Chalmers's 2018 own formulation of the problem intuitions in topic neutral terms include « We have special properties that are hard to explain », or « that are non-physical », « that provide special first-person knowledge », « that could be missing in robots », etc.¹¹. Obviously, none of these can be explained by the shared features of phenomenal concepts and yogi concepts, given the simple fact that thinking about brain states with yogi concepts do not give rise to these problem intuitions. To rule out objections like this, Chalmers suggests that the problem intuitions to be explained should feature concepts like *presence*, *acquaintance*, or *revelation*, etc. But while *presence* or *acquaintance* are not overtly phenomenal concepts, they are phenomenal nevertheless. They are constitutively tied to the concept of consciousness. We understand what it means to be *acquainted* with a mental state, or for a mental state to be *present*, in terms of our acquaintance with conscious states. It is hard to see how a topic-neutral explication of them could possibly go, and so how one might explain beliefs involving these concepts topic-neutrally, which undermines the debunking argument.

Take another, this time real world problem Chalmers discusses: the problem of why introspecting thoughts doesn't create the same problem intuitions as introspecting conscious experience. He thinks it is because the latter acquaints us with its subject, and the former doesn't. This is a version of the same problem we had in connection with beliefs involving yogi concepts. If acquaintance, as I maintain, doesn't have a topic-neutral explication, no topic-neutral description of those otherwise similar cognitive mechanisms, both plausibly involving direct concepts applied to mental states as they occur, could account for the fact that the one *acquaints* us with its object and the other doesn't. Only appealing to consciousness will make the right distinction between cases where the problem intuitions occur and similar cases where they don't. Formulating the problem intuitions topic-neutrally leads to bad explanations. And once we add notions such as *acquaintance* and *presence* to the mix, we can have good explanations, explanations that distinguish between the consciousness case and the yogi case, as well between thinking about consciousness and thinking about thought. But these explanations cannot be given in topic-neutral terms¹². The starting point of the debunking argument is ill-conceived.

¹¹ Chalmers 2018, p. 18.

¹² Chalmers considers this response to his meta-problem research program. As he puts it, « Some non-reductionists may embrace meta-problem nihilism : there is no solution to the meta-problem. Alternatively, if we understand the meta-problem more broadly as 'Explain our problem intuitions in topic-neutral terms, or explain why this is impossible', then the meta-problem nihilist says that any solution must take the second horn. » (p. 41).

Another way of understanding the problem with the debunking argument is that we only make topic-neutral judgments about consciousness, if at all, based on beliefs that feature phenomenal terms. Therefore, it is not surprising to find that we can only really explain the quasi-phenomenal judgements, that is, judgements about consciousness that do not appeal to phenomenal concepts, if we can also explain the problem intuitions *themselves*, stated in *phenomenal* terms. But of course, those problem intuitions, as we have already observed, cannot be explained in purely topic-neutral terms any more than consciousness itself can be explained in purely topic-neutral terms. In both cases, this is due to the conceptual isolation of phenomenal concepts from any physical or functional concepts a topic-neutral description might employ.

Conceding the point about explaining the problem intuitions *qua* problem intuitions stated in phenomenal language, or at least in the language of acquaintance, presence, etc., one might nevertheless suggest that, on the physicalist assumption, the problem intuitions can *also* be described in strictly neuro-scientific terms. This seems to count in favor of the debunking argument: we know that the problem intuitions – stated in the language of neuroscience – have a topic-neutral explanation, i.e., an explanation in the language of neuroscience. Does *this* interpretation make premise 1 of the debunking argument true? This is an interesting proposal¹³; I don't think, however, that it helps the debunking argument.

We are nowhere near knowing enough about the brain to identify any thought, including the problem intuitions, in neuroscientific terms, and it is highly uncertain when if ever it will be possible to do so. But – and this is the more important point – even if we had a full mapping of thought to neural functioning, we couldn't use any forthcoming explanation of them to debunk consciousness. The fact that these beliefs – conceptualized in the language of neuro-science – can be explained without explicit appeal to consciousness *as such* – i.e., without appeal to consciousness under phenomenal conceptualization – is a triviality. It does not prove that consciousness itself plays no explanatory role. In fact, it is to be *expected* that the explanation will invoke *consciousness in the language of neuroscience*. Simply leaving the language of consciousness, acquaintance, etc. behind in this way won't allow a proponent of illusionism to argue that consciousness or acquaintance does not figure at all in the explanation of the target phenomena. They are not implicated *as such* – i.e., in phenomenal language – when we describe the target phenomenal in neural terms, but that doesn't mean they are not nevertheless implicated in those neural terms, that the explanation doesn't *in fact* talk about consciousness in its neuro-physiological guise. To sum up, premise 1 is either demonstrably false or unsupported. This bodes ill for the illusionist project.

Granted, science and philosophy can and have gone against deeply held common sense views. Obvious examples concern the nature of physical objects (containing mostly empty space), the nature of causation (not an inner, unobservable force), or, more controversially, the nature of the self (not a mental substance) and free will (not incompatible with determinism). But the case of experience is not like that. In the case of experience, as I have argued above, the pressure that science and philosophy can bring to bear is nowhere near strong enough to justify doubt in one's own experience. There are no scientific discoveries incompatible with the existence of conscious experience; and there are no decisive philosophical arguments against it, much less a demonstrable incoherence in our concept of it. Consequently, it is not some overwhelming theoretical reason, accepted on balance and reluctantly in the face of the objections of common sense, that draws its adherents to it. What is it, then?

¹³ Thanks to François Kammerer for raising this point.

As far as I can see, the allure of illusionism consists in two key features. One is its very counter-intuitiveness. Some find the idea that our most deeply held views are a giant hoax exciting. But I think its most important attraction is adherence to what it mistakenly takes to be a hardnosed, thorough going scientific outlook. Illusionism appears to uphold the banner of science and reason against the prejudices of common sense. I suspect some scientifically minded philosophers find illusionism exciting in the way Darwinism was exciting in the 19th and early 20th century, unseating our cherished views about ourselves. They seem to think that illusionism is allied to science while those who insist on common sense about consciousness and introspection are mere reactionaries.

But there is another related aspect of illusionism that makes it especially worrisome. Illusionism fits well with *scientism*: the view that the best way to study everything, including matters of humanistic concern, is through science¹⁴. If consciousness doesn't exist, then humanistic studies appealing to experience in explaining history, art, or morality can't be of much value. Better to stick to cognitive neuroscience when studying behavior. This seems to be the view of some illusionists. Frankish (2016), for example, suggests that cognitive science should eliminate talk about phenomenal experience and replace it with talk about quasi-phenomenal properties. And this is certainly in the background of the views of some other notable illusionists, like Rey (1995) and Dennett (1988, 1991, 2017). Though Graziano (2016) and Dennett (1991) agrees that the illusion of phenomenal consciousness plays an important and evolutionarily explicable role in our mental lives, it is hard to imagine how a belief in its illusoriness would not undermine that role.

3. The belief in illusionism and its moral consequences

The humanities as well as everyday thought about human affairs appeal frequently to experience in their explanation of human behavior. The horror people feel upon exposed to gruesome sights, or the elation they experience at a mass rally, are conscious states with a lot of explanatory power. When we talk about what is meaningful in our own lives, we are often referring to a felt sense that certain activities are worthwhile. What we do in the wake of them *makes sense* in light of those experiences. If one supposes such experiences are illusory one will become dismissive towards such explanations. But the problem is even worse: being dismissive toward conscious experience makes one disinclined to think about oneself and other conscious beings in this introspective manner. Such an attitude would predispose one to see humans in the same way most of us see robots: unfeeling, meaningless (meat)-machines lacking moral status.

To illustrate the point, consider the following thought experiment. Suppose there was a super-intelligent organism – let's call her Zombie-Mary, with a nod to Frank Jackson's (1982) super-scientist Mary – that lacked any feeling or experience, a creature of pure thought. When she sees a roadside accident, she has no « gut reaction » to it: in addition to not experiencing color, sound, etc., she feels no aversion, no horror, no sadness, or, as the case might be, no morbid curiosity.

Zombie-Mary could know a tremendous amount about humans in biological, neuronal, and

¹⁴ See for example this statement from Alex Rosenberg : « The only solution to the problem faced by the humanities, history and (auto) biography, is to show that interpretation can somehow be grounded in neuroscience. That is job No. 1 for neurophilosophy. And the odds are against it. If this project doesn't work out, science will have to face plan B : treating the humanities the way we treat the arts, indispensable parts of human experience but not to be mistaken for contributions to knowledge. »

<https://archive.nytimes.com/opinionator.blogs.nytimes.com/2011/11/06/bodies-in-motion-an-exchange/>. For a counterargument see Ismael (2018).

information-processing terms¹⁵ – but she has never experienced the myriad ways in which something can be beautiful, painful, scary, or desirable. She doesn't know about these things from *the first-person perspective*, since she cannot think about them experientially. She says that she does – after all, she is a physical duplicate of Jackson's Mary – but according to our premise she is merely deluded into thinking she does.

I suggest it is clear that Zombie-Mary knows nothing of value, meaning, and human significance. Zombie-Mary does not only lack an understanding of pain, but also of the badness of pain. Nothing could be beautiful, or attractive, or horrifying to her, which makes her incapable of having moral concepts, as moral concepts are constitutively tied to experiential ones such as suffering or flourishing. We *make sense* of morality in terms of our experiential understanding of suffering and flourishing; and this is crucial for the way moral judgments motivate.

If illusionists are right, we are like Zombie-Mary (minus the super-human intelligence), rendering *our* moral concepts illusory as well, another counterintuitive consequence of illusionism. At the same time, it would mean that we are also not *proper moral subjects*, lacking the ability to suffer or flourish in an experiential sense.. This is clearly wrong; a barely coherent position. Illusionists might retort that value only depends on the existence of desires and goals understood as dispositions, that value doesn't depend on conscious experience, and so there is no problem accounting for moral worth even though conscious experience doesn't exist. I disagree. This would be a reinterpretation of our concept of value. We cannot understand value in the purely third-person language of function and disposition any more than we can understand experience in the purely third person language of function and representation.

The situation is different than it had been prior to the twentieth century. Everyone used to take it for granted that it is through lived experience that we primarily relate to the world and the good or bad in it, and it wasn't controversial that awareness of experience is a crucial component of self-knowledge, and through empathy and imagination, of knowledge of other people. Now we are able to think about ourselves not only as conscious subjects but also as information processors, or a collection of neurons firing away in our skulls. We are all of that – yet this perspective can't replace our subjective view of ourselves as conscious beings, it can't be the full story we are telling ourselves, *even if conscious experience is a purely physical or functional phenomenon*¹⁶. A purely third-person scientific perspective leaves out most of what makes it possible to understand each other, and to chart our course in the world¹⁷. We are creatures who can make sense of ourselves and others primarily in phenomenal terms.

None of this requires an abandonment of physicalism. The key to physicalist consciousness realism is that there are two, radically different ways to conceptualize consciousness. Consciousness can be studied scientifically and – at least in principle – we can come to discover its neural nature. But we can also form direct, introspective concepts of it. We have this ability only regarding certain special neural states. That it is important that we engage these neural states via our introspective phenomenal concepts, doesn't imply that there is anything wrong with physicalism. We can do different things with

¹⁵ If you are doubtful that a zombie can have intentional states, imagine instead PartialZombie-Mary who experiences the usual perceptual properties and sensations but has no affective experience at all (could not experience something as beautiful or scary, etc.).

¹⁶ As I have argued above, the phenomenal concept strategy has demonstrated that there is no contradiction between the metaphysical claim that all phenomena are purely physical and the conceptual observation that we can't explicate phenomenal concepts in terms of functional, representational, or physical concepts.

¹⁷ See Ismael (2018)

water when we are equipped with the concept 'water' that we couldn't do if we could only think about it in terms of H₂O or some massively complex fundamental physical property. Maybe this limitation doesn't apply to angels, but it does to humans. And similarly, even if conscious experience is entirely physical, we need to engage it introspectively to learn about the value of people and things, to empathize and understand others, etc.

Illusionism, because of its implications for value and moral worth, poses a problem that is not merely theoretical. A belief in illusionism is not simply wrong, it leads to morally detrimental consequences. This connection seems to have some empirical corroboration. For example, in a study on mind perception¹⁸, the authors have found that people want to avoid harming other creatures to the degree that they attribute to them the capacity for experience. This suggests that illusionist beliefs might lead to less empathy and care for oneself and others. If one doesn't believe in moral worth, one will stop caring.

Metaphysical beliefs of every sort have practical consequences. For example, recent studies have found a correlation between a belief in the soul on the one hand, and helping behavior (Genschow 2023) and a sense of meaning and wellbeing in the world (Timmermann 2021) on the other¹⁹. But of course, I am not advocating for the soul view on the grounds of its practical superiority; its practical consequences can be legitimately trumped by the imperative of forming an accurate view of the world. We cannot unlearn what we have learned philosophically and scientifically, in order to hold on to a more satisfying, pre-modern view of ourselves, even if denying the existence of the soul has bad practical consequences. Similarly, if illusionism were true, we would have to take the bitter pill – the destruction of what we took to be our moral concerns – for the sake of truth. I believe, however, that illusionism is deeply wrong on the facts. Even though the soul doesn't exist, conscious experience exists, and it is central to a human life. The belief in consciousness and an inner life is the last redoubt of our pre-modern view of ourselves but it is a safe harbor: there are simply no good reasons to give it up. Illusionism is based on arguments that can be effectively rebutted. That it is also wrong practically adds a further, and more urgent, reason to resist illusionism's influence on philosophy, cognitive science, artificial intelligence, and culture at large.

Bibliography

- Baars B., *A Cognitive Theory of Consciousness*, Cambridge, Cambridge University Press, 1988.
- Bealer G., « Mental Properties », *Journal of Philosophy*, vol. 91, 1994, p. 185–208.
- Block N., « Max Black's Objection to Mind-Body Identity » in D. Zimmerman (ed.), *Oxford Studies in Metaphysics*, II, Oxford, Oxford University Press, 2006, p. 3-78.
- Chalmers D., « The Content and Epistemology of Phenomenal Belief », in Q. Smith and A. Jokic (eds.), *Consciousness: New Philosophical Perspectives*, Oxford, Oxford University Press, 2003, p. 220-73.
- Chalmers D., « Phenomenal Concepts and the Explanatory Gap », in T. Alter & S. Walter (eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, Oxford, Oxford University Press, 2007.
- Chalmers D., « The Two-Dimensional Argument Against Materialism », in B. McLaughlin (ed.), *The*

¹⁸ Gray, H.M ; Gray, K. ; Wegner D. M. (2007).

¹⁹ Though these studies don't track down the possible ways in which a belief in the soul might lead to these effects, one might hypothesize that it is the lack of illusionist beliefs, illusionism being incompatible with the soul view, that is positively correlated with helping and psychological well-being. The soul view's most popular alternative, physicalism – misguidedly, but with increasing success – invites an illusionist self-view that might lead to these negative effects.

- Oxford Handbook of the Philosophy of Mind*, Oxford, Oxford University Press, 2009.
- Chalmers D., « The Meta-Problem of Consciousness », *Journal of Consciousness Studies*, vol. 25, n°9–10, 2018, p. 6–61.
- Dennett D. C., « Quining Qualia », in A. Marcel & E. Bisiach (eds.), *Consciousness in Modern Science*, Oxford, Oxford University Press, 1988. Reprinted in W. Lycan (ed.), *Mind and Cognition : A Reader*, Cambridge, MIT Press, 1990 ; A. Goldman (ed.), *Readings in Philosophy and Cognitive Science*, Cambridge, MIT Press, 1993.
- Dennett D. C., *Consciousness Explained*, New York, Little, Brown, 1991.
- Dennett D. C., *From Bacteria to Bach and Back: The Evolution of Minds*, New York/London, W.W. Norton & Company, 2017.
- Duncan M., « Acquaintance ». *Philosophy Compass*, vol. 16, n°3, 2021, p. 1–19.
- Frankish K., « Quining diet qualia », *Consciousness and Cognition*, vol. 21, n°2, 2012, p. 667–676.
- Frankish K., « Illusionism as a Theory of Consciousness », *Journal of Consciousness Studies*, vol. 23, n°11–12, 2016, p. 11–39.
- Giustina A., « Introspective knowledge by acquaintance », *Synthese*, vol. 200, n°2, 2022, p. 1–23.
- Gray H.M., Gray K., Wegner D. M., « Dimensions of Mind Perception », *Science*, vol. 315, n°5812, 2007, p. 619.
- Graziano M., « Consciousness engineered », *Journal of Consciousness Studies*, vol. 23, n°11–12, 2016, p. 116–123.
- Graziano M., *Rethinking Consciousness: A Scientific Theory of Subjective Experience*, New York, W. W. Norton & Company, 2019.
- Genschow O., « It Is Belief in Dualism, and Not Free Will, That Best Predicts Helping: A Conceptual Replication and Extension of Baumeister et al. (2009) », *Personality and Social Psychology Bulletin*, 2023, p. 1–12.
- Ismael J., « Why (study) the Humanities? », in S. Grimm (ed.), *Making Sense of the World: New Essays on the Philosophy of Understanding*, Oxford, Oxford University Press, 2018.
- Jackson F., « Epiphenomenal Qualia », *Philosophical Quarterly*, vol. 32, 1982, p. 127–36.
- Kammerer F., « The hardest aspect of the illusion problem – and how to solve it », *Journal of Consciousness Studies*, vol. 23, n°11–12, 2016, p. 124–139.
- Kammerer F., « The illusion of conscious experience », *Synthese*, vol. 198, 2021, p. 845–866. <https://doi.org/10.1007/s11229-018-02071-y>
- Kirsch A., (2023). *The Revolt Against Humanity: Imagining a Future Without Us*, New York, Columbia Global Reports, 2023.
- Kripke S., *Naming and Necessity*, Cambridge, Harvard University Press, 1972.
- Levine J., *Purple Haze: The Puzzle of Consciousness*, Oxford, Oxford University Press, 2001.
- Levine J., « Phenomenal concepts and the materialist constraint », in T. Alter & S. Walter (eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and physicalism*, Oxford, Oxford University Press, 2007, p. 145–165.
- Loar B., « Phenomenal States », *Philosophical Perspectives*, vol. 4, 1990, p. 81–108.
- Loar B., « Phenomenal States: Second Version », in N. Block, O. Flanagan & G. Guzeldier (eds.), *The Nature of Consciousness: Philosophical Debates*, Cambridge, MIT Press, 1997, p. 597–616.
- Nagel T., « What Is It Like to Be a Bat? », *Philosophical Review*, vol. 83, 1974, p. 435–50.
- Nida-Rümelin M., « Grasping Phenomenal Properties », in T. Alter & S. Walter (eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and physicalism*, Oxford, Oxford University Press, 2007.
- Papineau D., « The rise of physicalism », in C. Gillett & B. M. Loewer (eds.), *Physicalism and its Discontents*, Cambridge, Cambridge University Press, 2001.

- Papineau D., *Thinking about consciousness*, Oxford, Oxford University Press, 2002.
- Peacocke C., « Descartes defended », *Proceedings of the Aristotelian Society*, vol. 112, n°3, 2012, p. 109–125.
- Pereboom D., *Consciousness and the Prospects of Physicalism*, Oxford, Oxford University Press, 2011.
- Pereboom D., « Russellian Monism, Introspective Inaccuracy, and the Illusion Meta-Problem of Consciousness », *Journal of Consciousness Studies*, vol. 26, n°9–10, 2019.
- Robinson H., « The Anti-materialist Strategy and the Knowledge Argument », in H. Robinson (ed.), *Objections to physicalism*, Oxford, Oxford University Press, 1993.
- Rey G., « Towards a projectivist account of conscious experience », in T. Metzinger (ed.), *Conscious Experience*, Thorverton, Imprint Academic, 1995, p. 123–42.
- Rosenthal D., « Explaining consciousness », in D. J. Chalmers (ed.), *Philosophy of Mind: Classical and Contemporary Readings*, Oxford, Oxford University Press, 2002.
- Stoljar D., « Physicalism and phenomenal concepts », *Mind and Language*, vol. 20, n°2, 2005, p. 296–302.
- Tononi G., « Integrated information theory », in M. Velmans & S. Schneider (eds.), *The Blackwell Companion to Consciousness*, Oxford, Blackwell, 2007.
- Schwitzgebel E., « Phenomenal Consciousness, Defined and Defended as Innocently as I Can Manage », *Journal of Consciousness Studies*, vol. 23, n°11–12, 2016.
- Schwitzgebel E., « Introspection », *Stanford Encyclopedia of Philosophy*, 2019. <https://plato.stanford.edu/archives/win2019/entries/introspection/>
- Timmermann C. et al., « Psychedelics alter metaphysical beliefs », *Nature*, vol. 11, n°1, 2021.
- White S., « The Argument for the Semantic Premise », in T. Alter & S. Walter (eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and physicalism*, Oxford, Oxford University Press, 2007.