

The Phenomenological Negation of the Causal Closure of the Physical

John Thornton

The Institute for Integrated and Intelligent Systems, Griffith University, Australia
j.thornton@griffith.edu.au

1 Preliminaries

The aim of this paper is to provide a phenomenological demonstration that the causal closure of the physical is false. To begin I should like to clarify the notion of the physical I have in mind. As an example, consider the experience you are having now. I am assuming your eyes are open and you are seeing something. That something, taken in the ordinary sense, is the physical world, made up of physical spaces and objects, such as tables and chairs, and physical events involving such spaces and objects. In everyday conversation we know quite well what it means for something to be physical, and we can reliably indicate whether something is, in fact, physical. The concept is thereby used to distinguish between things and events that exist (in a way that can be publicly demonstrated), and the various experiences that can only be known to us privately or inwardly. It is this notion of the physical in contradistinction to entirely private, phenomenal, subjective experience that I shall be considering in this paper.

A paradigm example of an entirely subjective experience is my acquaintance with sensory colour, such as the redness of a particular shade of red. The demonstration that such experience is entirely subjective is my inability to express or indicate to anyone else what my experience of the particular red colour quality *qua* quality is (in itself). I can indicate the object whose objective red colour is known to me on the basis of the colour quality. I can indicate the photons that are emitted from the object, the responses that occur in my retina, and the events occurring in my brain. But at no point in that chain of events can I indicate directly *to you* what it is that I am experiencing as the quality or essence of the colour of the object.

The best I can do is to ask you to inspect your entirely subjective experience of the colour of the object I am looking at, and work on the assumption that the colour quality you are experiencing is *similar* to the one I am experiencing (the likelihood that our qualitative experiences should concur exactly, given the differences in the responses of our respective visual systems, is not high). Perhaps, with a suitable brain-o-meter, we could observe that events of the same class are occurring in both our brains and, given that membership of this class determines the colour quality that is experienced, we could conclude we are both experiencing the same quality. The trouble here is that there is no conceivable experiment that could confirm, in general, that such and

such a class of brain events determines such and such a colour experience (because my experience of colour quality is entirely subjective and so cannot be objectively corroborated).

Of course, I can simply decree that when a brain state of such and such a class obtains, then such and such a quality is experienced. The advantage (and disadvantage) of such a decree is that while no one can confirm it, no one can deny it either. I can even express my decree in the form of a psycho-physical law and claim a scientific provenance, because (after all), all scientific explanation must ground out in something one cannot explain, some ultimate brute fact or relation. Why not the brute fact of psycho-physical law? The problem here is that what we usually accept as scientific brute facts are facts whose effects we can observe objectively. So, for example, if we take Schrödinger's wave equation to be a brute fact, this is a brute fact whose validity can be tested by experimental observation. In contrast, the psycho-physical law that such and such a class of brain states determines such and such a qualitative experience cannot *in principle* be tested by experiment. What am I to say? "Yes I am experiencing quality x now"? But, as Wittgenstein pointed out, x cannot be defined. I have no precise idea what you mean by x and you have no precise idea what I mean by x . I cannot even verify that you are experiencing what-I-call-colour and not something that I would find incomprehensibly unfamiliar. I can only see *my* beetle, the one you can never see.¹ If we are to take Popper's maxim that a theory is only scientific if it can be falsified, then the notion of psycho-physical law is *not* scientific.² Its function is rather to uphold a certain metaphysical position concerning the determinability of phenomenal experience.

So, for the purposes of this paper, the physical is the non-phenomenal, and the phenomenal is the pure quality of subjectivity, that which can only be known on the basis of direct or immediate experience. Given this category of the physical, we can now define the causal closure of the physical as the principle that 'every physical event is determined, in so far as it is determined at all, by preceding physical conditions and laws' (Montero & Papineau 2005, p. 233). Causal closure therefore denies any *independent* causal efficacy to phenomenal experience. Accordingly, my experience of the quality of redness can enjoy a kind of epiphenomenal existence, but it can in no way, in and of itself, cause another physical event to occur. If it appears that such events do occur, such as my saying that "I am having a phenomenal red experience now," it must be that some underlying physical state has determined that utterance, just as some related physical state has determined my experience of the red quality.

In the remainder of the paper, I shall attempt to show that this notion of causal closure is false. The aim is to arrive at the insight that causal closure is false by means of a *direct seeing* that contains its own self-evidence. How this is to be achieved will become clearer as we continue.

¹ Wittgenstein's point was that sensations (phenomenal qualities or beetles in boxes) cannot be meaningfully referred to as thing-objects within a standard language-game (Wittgenstein 2001, p. 85, Section 293).

² See (Popper 1959/2002).

2 The Paradox of Phenomenal Judgment

I think the most comprehensive and influential defence of the principle of causal closure is that provided by David Chalmers in his account of the paradox of phenomenal judgment.³ The paradox arises from having already accepted the fundamental premise that reality is ontologically divided into the objectively physical and the subjectively phenomenal on the basis that the objectively physical is causally closed. Given this division, the ability to make correct judgments concerning subjectively phenomenal experiences appears paradoxical because phenomenal judgments are expressed as physical events, and yet the experiences of phenomenal quality about which we judge are not supposed to have any independent effects on physical events. Considered counterfactually, this amounts to supposing that speech acts concerning phenomenal experience will unfold in just the same way, *whether or not* there is any accompanying phenomenal experience.

Chalmers' answer to the paradox is to introduce the notion of *pure phenomenal concepts*. These concepts are physically instantiated in the brain, but bear no reference or relation to any objectively physical entity. Instead they refer to pure phenomenal qualities that are demonstrated directly in an immediate phenomenal experience. For example, I could be looking at a green leaf on a plant in front of me. Firstly, I grasp my experience of a patch of uniform greenness on the leaf as something distinct from my experience of the objectively physical leaf. I can indicate the experience as *this* green quality, meaning the colour quality I am experiencing now, subjectively, and only for as long as I am paying attention to the actual quality of that colour. An objectively physical account of colour perception would associate this subjective experience with certain processes occurring in my objectively physical visual system. However, the pure phenomenal concept does not refer to these physical processes either. It refers only to the experience itself, as a phenomenal experience.

Chalmers' argument concerning the paradox of phenomenal judgment contends that my consciousness of phenomenal quality makes no difference to the physical functioning of my brain or my speech behaviour. His idea is that the physical events in my brain that correspond to my forming a pure phenomenal concept, and uttering a judgment employing that concept, are determined (as far as they are determined), according to causal closure, i.e. by the preceding physical events and the physical laws that govern them. My experience of phenomenal quality is something additional that accompanies the physical formation of a pure phenomenal concept, and *constitutes* the content of that concept. The pure phenomenal content of the concept does not *cause* the concept to be formed, and thereby does not violate the principle of causal closure.

³ The defence of causal closure is first laid out in Chapter 5 of *The Conscious Mind* (Chalmers 1996) and then elaborated in subsequent papers that are now amalgamated in Chapters 8 and 9 of *The Character of Consciousness* (Chalmers 2010a, Chalmers 2010b).

3 The Acquisition of Phenomenal Concepts

The problem with Chalmers' account is that it fails to resolve the paradox of phenomenal judgment. It rather shows how it is possible for a physically determined, unconscious entity to mimic a certain kind of human behaviour. What is not addressed is how such an unconscious entity could acquire the ability to wield pure phenomenal concepts in the first place. Instead, we are introduced to a fully formed brain, one that already possesses such abilities, and we are shown how this brain may continue to function in the absence of consciousness, in such a way as to utter judgments concerning pure phenomenal concepts that would satisfy a Turing test.

However, Chalmers offers an account of how an unconscious entity could acquire phenomenal concepts in *The Conscious Mind*. There he uses the notion of an information space and a processing system that finds itself within that information space, to explain how such a system could become puzzled about its experience:

The crucial feature here is that when the system perceives a red object, central processes do not have direct access to the object itself, and they do not have direct access to the physical processes underlying perception. *All* that these processes have access to is the color information itself, which is merely a location in a three-dimensional information space.

... Indeed, as far as central processing is concerned, it simply *finds itself* in a location in this space. The system is able to make distinctions, and it *knows* it is able to make distinctions, but it has no idea how it does it. We would expect after a while that it could come to *label* the various locations it is thrown into – “red,” “green,” and the like – and that it would be able to know just which state it is in at a given time. But when asked just *how* it knows, there is nothing it can say, over and above “I just know, directly.” If one asks it, “What is the difference between these states?” it has no answer to give beyond “They’re just different,” or “This is one of *those*,” or “This one is *red* and that one is *green*.” When pressed as to what that means, the system has nothing left to say but “They’re just different, qualitatively.”

... Given this kind of direct access to information states, then, it is natural to expect the system to use the language of “experience” and “quality” to describe its own cognitive point of view on perception. And it is unsurprising that all this will seem quite strange to the system: these immediately known, ineffable states, which seem so central to its access to the world but which are so hard to pin down. Indeed, it is natural to suppose that this would seem odd to the system in the same sort of way that consciousness seems odd to us.

So this is the beginning of a potential reductive explanation of our judgments about consciousness: these judgments arise because our processing system is thrust into locations in information space, with direct access to those locations but to nothing else. The direct knowledge will strike the system as a brute

“quality”: it knows that the states are different, but cannot articulate this beyond saying, in effect, “one of *those*.” This immediate access to brute differences leads to judgments about the mysterious primitive nature of these qualities, about the impossibility of explicating them in more basic terms, and to many of the other judgments that we often make about conscious experience. (Chalmers 1996, p. 290-291).

It is to be noted here that Chalmers gives no account of the acquisition of *pure* phenomenal concepts. All that is shown is how unconscious systems can come to report and reason about the various physical configurations in which they find themselves. Chalmers’ crucial assumption is that the central processes in his unconscious system ‘do not have direct access to the physical processes underlying perception.’ The ‘mystery’ for this system does not concern the existence of pure phenomenal qualities – if it is unconscious we must assume that it has no experience of such qualities. The mystery is rather a consequence of the system not ‘understanding’ the principles of its own operation. This kind of mystery is easily cleared up. The system only needs access to information concerning the design, construction and operation of its own components. Like a disciple of Dennett and the Churchlands, it will then ‘understand’ that its previous talk of mysterious qualities was an illusion. It only *seemed* that way because it did not have enough information concerning the physical realisation of its information space. Given this information, the entity will immediately ‘understand’ that its ‘experience’ of red can be entirely explained in terms of physical changes occurring within its physical components that cause it to utter statements to the effect that something is red when it is placed in front of paradigmatically red objects. In Chalmers’ terminology, the system’s phenomenal concepts will all be *relational* phenomenal concepts. They are relational because the corresponding phenomenal qualities are determined in relation to something else, i.e. states of affairs in the physical world that act as causes both in the forming and deployment of the relational phenomenal concepts. There will be no question of an additional ‘quality’ that is ‘experienced,’ there will just be the dispositions to utter certain phrases in certain situations concerning the detection of certain physical stimuli.

The question that Chalmers’ paradox of phenomenal judgment poses is how an unconscious entity could acquire a *pure* phenomenal concept, i.e. a concept that refers to a pure phenomenal quality that bears no relation with any physical state, property or process. We, as conscious entities, already have sufficient information concerning the operation of the physical brain to see that our phenomenal experiences have corresponding physical manifestations. But this knowledge has not caused us to discard our notion that there are experiential qualities that are not captured by physical descriptions of the functioning of the brain. The task for Chalmers, in order to resolve the paradox of phenomenal judgment, while maintaining causal closure, is to explain how an unconscious entity, entirely on the basis of its own ‘experience,’ and knowing all the details of its own physical operation, could come to the conclusion, like us, that there

is something more to being conscious than it already knows on the basis of physical science.

4 The Pre-Understanding of Pure Phenomenal Quality

Chalmers' account assumes that a pure phenomenal concept spontaneously comes into play once I direct my attention onto a pure phenomenal quality that is immediately present in my phenomenal experience and attempt to form a direct phenomenal belief concerning that phenomenal quality. However, in order to direct my attention in this way, I must already have an idea of what it is that I will find (i.e. an intention, or directedness-toward). That is, I must already understand that there is such a 'thing' as pure phenomenal quality that corresponds to my pure phenomenal concept.

It is this *pre*-understanding that Chalmers takes for granted. In order for his account to work, he needs to explain how it is that an unconscious entity could come to *notice* that there is such a 'thing' as a non-relational, pure phenomenal quality, in the first place. And this is something that an unconscious entity cannot do, because the very thing that the unconscious entity is unconscious of is non-relational, pure phenomenal quality. *If* an unconscious entity were to report having an experience of pure phenomenal quality, we would have to conclude that there was some error in its construction, an error that could be traced to some malfunctioning physical structure, and consequently rectified.⁴

Chalmers attempted to show how just such an error could be made in his account of information spaces quoted earlier. However, the information space scenario assumed that the unconscious entity remains in ignorance concerning the physical realisation of its information space. Once that artificial barrier is removed, i.e. once the system is given the same access to the world as we have, then Chalmers' argument collapses. It collapses because the rational response of such a system, once in possession of all the relevant information, is to assert, with Dennett, that eliminative physicalism is true.

The problem for eliminative physicalism, as Strawson has so trenchantly pointed out,⁵ is that if we know anything at all, we *at least* know there is phenomenal experience of pure phenomenal quality. The question is, how can we know this, if phenomenal experience has no independent causal effect on our behaviour? Again, Dennett is right: if phenomenal experience has no independent causal effect on our behaviour, then we can't know that there is such a thing as phenomenal experience. Therefore our idea of phenomenal experience is an illusion.⁶

Chalmers' argument does not touch this conclusion. The problem is not to explain how I can refer to phenomenal quality using a pure phenomenal concept in such a way

⁴ See (Elitzur 2009) for a refutation of causal closure that shows why an argument that rests on the premise of a machine that falsely concludes it is conscious must fail.

⁵ See (Strawson 2008).

⁶ See (Dennett 1991, pp. 369-411).

that the phenomenal quality is not causally implicated in the explanation. The problem is to explain how it is that I am able to maintain the *conviction* that there is such a thing as pure phenomenal experience in the face of the physical evidence concerning the operation of my brain. It is only on the basis of this conviction that I am able to distinguish between a pure and a relational phenomenal concept in the first place. For example, let us assume I am in possession of a relational phenomenal concept that is formed as I demonstrate a particular shade of green to myself, e.g. “*this* shade of green.” My concept refers to the phenomenal experience occurring in me as a result of looking at a green leaf in front of me and so is *related* to my act of demonstration. It is quite conceivable that Chalmers’ unconscious system could form a structure within itself that corresponds to this concept. When questioned further it would identify that phenomenal experience with certain events that the light emitted from the leaf causes to occur within its components, just as I can refer my phenomenal experience to certain events that the light causes to occur in my brain. The crucial difference comes when I ask the system to distinguish between its phenomenal experience of green and the corresponding activity of its components. It will have to say there that they are one and the same thing. But what about me? What is it that allows me think that my phenomenal colour experience is something more than the activity occurring in my brain? If my phenomenal experience is causally determined by the operation of physical law, then that experience cannot reach out of its causal dependency and start independently influencing my neurons. I may be passively conscious of the greenness of my phenomenal experience, but, according to causal closure, that passive consciousness can have no independent effect on the operation of my brain. So, despite my having the experience, I will be unable to form any thought that corresponds to or even registers the experience as anything more than the physical realisation of that experience. I will be necessarily mute on the subject, not just outwardly, I will be unable to even *think* that there is anything more to my experience than the physical functioning of my brain. I will be trapped in my relational phenomenal concepts, having my pure phenomenal experience, but powerless to form concepts with which to even indicate that such experience is occurring.

And yet, when I contemplate my phenomenal experience of the greenness of the leaf, I do form the conviction that there is something more to the experience than can be explained in objectively physical terms. If I examine this carefully, I do not form the conviction on the basis of some form of induction from experience. I know immediately, from within the experience itself. I cannot offer any *positive* justification, I immediately *see* that this is the case.

5 Phenomenal Concepts and their Reference

In the context of Chalmers’ account, phenomenological ‘seeing’ is the act whereby I become aware of the pure phenomenal content of a pure phenomenal concept. This

awareness occurs while I am attending to the colour of a suitably coloured object and ceases as soon as I stop so attending. The concept itself is unusual. Firstly, it only ‘exists’ for as long as I am engaging in a correct demonstration, and secondly, it refers to something peculiar: a pure phenomenal quality. It is this pure phenomenal quality that is (supposedly) ‘seen.’

To examine this further, we shall require a model of what it means for something to be a concept, and what it means for that concept to refer to something. I shall take it that in ordinary experience, we encounter the world on the basis of a certain *perceptual intentionality*. That intentionality structures our present experience of the world both according to the actual data streaming through our sensory receptors and according to our past experience (insofar as we recognise the world as something familiar). For example, if I consider a perception of a coloured object in front of me now, I can say that my perceptual intentionality intends the object as a physical object, located in physical space, having surfaces coloured in such and such a way, and that the data streaming through my sensory receptors is confirming that intentionality. In addition, my recognising the object as, for instance, a pen, and my classifying the pen as being blue, involves certain pre-existing abilities that distinguish pens from pencils, blue things from red things, and so on. I take the possession of such abilities to indicate the presence of a *concept* of the thing or property that is successfully distinguished.

The pen itself, I take to be associated with a certain *intentional object*. This intentional object is what I am referring to when I think and speak of the pen. If, in fact, I am perceiving a physical pen that corresponds to this intentional object, then that physical pen is *identical* to the intentional object. My perceptual intending of the pen can be thought of as an expectation (instantiated by physical processes in my brain) of a certain abstract form (instantiated within the stream of data arriving from the sensory receptors). If that form is detected, then I experience the identity of the form of my intending with a form of the world, i.e. I experience the conviction that the thing I am intending is really ‘there’ in the world.

However, I can also think and speak of things that do not exist in the way a physical pen exists (such as Sherlock Holmes and Doctor Watson) and I can *imaginatively* refer to such intentional objects, or I can refer to things as having been in the past (in memory), or I can expect the appearance of things in the future (in anticipation), or I can refer to entities I hypothesise may exist, and so on. In addition, I can intend or refer to *properties* of intentional objects, such as my pen being blue, and to *classes* of intentional object that share a common set of properties, such as blue pens, and to *states of affairs* involving relations between groups of intentional objects, such as the blue pen being on the table, and to *events* that involve changes in intentional objects, such as the blue pen falling off the table, and so on.

In speaking and thinking of these various kinds of intentional object, and of the properties, classes and relations that obtain between them, I am employing *concepts*. According to this model, it is the concept that refers, and its reference, in the first

instance, is to an intentional object, or to the properties, classes and relations that obtain between such objects. As already mentioned, if a particular intentionality corresponds with the world (in a way that is inter-subjectively agreed within a community of language users) then any correctly formed concept that refers to that intentionality, *ipso facto* refers to the world itself. Finally, I take it that my intending of intentional objects (and the properties, classes and relations that obtain between them), and my use of concepts that refer to these intentional acts, are associated with and depend upon (i.e. supervene on) certain physical processes occurring in my brain.

The point here is not to argue whether such an account of reference is going to work in all cases, it is to provide a framework within which we can make clear the kinds of problems that are involved in trying to speak of phenomenal quality and how these problems are covered over if we do not remain vigilant in the way we use language and in the way we demonstrate or gain access to pure phenomenal experience.

Firstly, it is necessary to examine how colour concepts work in ordinary language. For example, how would you demonstrate to yourself the quality of the colour blue? Perhaps you would look around and point to the first blue object you could find and say (inwardly) “*this* is blue.” Or perhaps there are no blue objects present, so instead you remember a blue object, or you imagine a *patch* of blue. In each case you must look at, or think of, or imagine ‘something,’ i.e. you have to form an intention, and having formed the intention, you require the presence of a corresponding intentional object in order to have the fulfilment of a present experience of blueness. Even the imagination of a pure field of blue colour is the imagination of an intentional object: it is intended *as* something imagined, it has a certain imaginary extension, it either has borders, or it fades into indistinctness. Whereas blueness itself, as a pure phenomenal quality, has no spatial delineation. Similarly, if I were to experience a *sensation* of blue, then the sensation would be *another kind* of intentional object. For example, I can produce the afterimage of a colour by closing my eyes after looking at a bright light. Such an experience can be intended as a sensation, i.e. as the event that appears to be sensating in my eyes (like the tingling sensation that can be experienced in the hands). In the case of the afterimage, my intentional object is the afterimage sensation, and it is the *sensation* that has the property of being coloured, just as, when I look at the pen, it is the *pen* that has the property of being blue.

In these examples, the *quality* of blue is made present via the intending of an intentional object that has the (objective) *property* of being blue. The fact of an intentional object’s being blue is a relational property that depends on the object and brings into play our ordinary, relational, public concept of colour. This concept is relational because the colour reference can be distinguished (in principle) in relation to the configuration of the physical world. So, for example, when I imagine a blue pen, certain areas of my brain become active that also become active when I perceive a blue physical object. And so on. In short, the property of something’s being called or distinguished as blue is determined by the physics, and our correct use of the normal concept of blue can

finally only be explained in terms of a set of suitably described physical entities and behaviours. The same goes for our ability to make fine distinctions between different shades of colour.

We shall call this concept of colour an *objective relational* concept because it refers to colour as a relational property of an intentional object that can be explicated physically. In contrast, when we consider a pure phenomenal concept, we are unable to give a physical account of how it is formed and to what it refers. To make this clear, consider again the case of my looking at a blue pen. If we follow the physical story, we can see how the presence of the pen activates an objective relational concept that enables me to conceive that there is a blue pen in front of me now. Following Chalmers, I can form the intention to highlight a particular uniform shade of colour on the surface of the pen, and form a concept that enables me to indicate the colour as *this* colour. Such an action can still be understood in terms of my forming an objective relational concept that refers to a property of an intentional object. The feature that has gone unmentioned, is that the demonstration of the colour of a certain area on the surface of the pen is accompanied by an experience of that colour having a certain quality of blueness. It is this quality, that, so to speak, hovers above the physicality of the pen, the photons, and the physical events occurring in my brain, and that remains unaccounted for in the physical account of the reference of my objective relational concept.

6 Phenomenological ‘Seeing’

The point of the preceding discussion is to understand that in using a pure phenomenal concept, and in thinking generally about such concepts and their referents, we are not dealing with concepts as they are usually understood and employed. Normal concepts (objective relational concepts) have a reference that can be fixed objectively in relation to the physical world. A pure phenomenal concept has no such reference. That means, if you employ a pure phenomenal concept as if it were a normal objective relational concept, you misuse it, and speak a form of nonsense. As Chalmers has already described, employing a pure phenomenal concept requires a *bringing to presence* of the actual quality. Unless you are able to *explicitly* demonstrate a *particular* pure phenomenal quality to yourself in this way, your *general* use of the concept will remain ungrounded. It is not enough to defer to another’s use of the concept, e.g. by saying when I refer to a pure phenomenal quality, I am referring to what you are referring to. This works for objective relational concepts, because there is an objective fact of the matter that can (in principle) be verified. Whereas, when I refer to a pure phenomenal quality, I am referring precisely to that subjective quality known immediately to me. Whether you also experience the same quality is something I cannot verify. To repeat: if you are unable to bring such an experience of pure phenomenal quality to explicit consciousness, then you are also unable to correctly use any language that refers to pure phenomenal quality.

This is not a minor issue. It is clear that *many* philosophers who publish papers discussing phenomenal experience have been unable to bring an experience of pure phenomenal quality to explicit consciousness. Instead, the experience remains implicit, covered over by a sophisticated preunderstanding that determines the supposed referent of a pure phenomenal concept in advance. In order to achieve an *explicit* experience of phenomenal quality, one has to suspend, or bracket, any concept that purports to already understand what it is that is experienced. This is not a matter of putting aside one concept and replacing it with another. As we have already demonstrated, a pure phenomenal quality cannot be represented by means of an objective relational concept, no matter how sophisticated. It is just not that kind of ‘thing,’ i.e. it is not a property of an intentional object.

This brings us to the phenomenological demonstration itself, i.e. the enactment of a phenomenological ‘seeing’ of pure phenomenal quality. Such explicit demonstration requires the creation of a discontinuity in the normal, more or less continuous process of intending the intentional objectivity that constitutes our normal waking experience of being in the world. This can be considered as a kind of limited or partial phenomenological epochē, enacted specifically in relation to a demonstration of pure phenomenal quality.⁷ What is suspended is the everyday conceptualising intentionality that prefigures what you are about to experience on the basis of what you have previously experienced. Such prefiguring is habitual, natural and necessary. However, it is only on the basis of abstaining from such conceptualisation that the essence of colour as something non-conceptual can emerge. Otherwise the understanding remains bound to the intentional objectivities that the qualities exhibit, and exhibit in such a way that the constituting function of phenomenal quality within consciousness remains covered over. This is not a matter of positively attempting to conceive of phenomenal quality. The *fact* is that the pure phenomenal quality is already present, and not in a hidden way. The only obstacle to encountering this quality *directly* is having already understood the experience according an objective relational conceptuality (one can only repeat this in different ways).

Despite the preamble, what is being indicated is essentially simple. You find a coloured object in front of you. You look at the colour of the object. You stop interpreting your experience of this colour as being the colour of the object in front of you, or as being the colour of any ‘thing’. You allow the colour to emerge as the colour that it is. You look at the colour experience itself, its quality, its essence. You look at what it is that makes your experience of blue *blue*. You remember that these words are indications only, that we cannot refer to this experience using ordinary objective rela-

⁷ The phenomenological epochē intended here is related to Husserl’s epochē of the objective sciences (Husserl 1970/1992, pp. 135-137), except we are bracketing the objective relational conceptualisation of colour only, rather than bracketing ‘all objective theoretical interests.’ However, in practice, such a bracketing of colour experience also requires a bracketing of all objective scientific understandings that would, from the outset, deny the possibility of a direct intuition of phenomenal colour.

tional concepts. And yet, it is only once the pure phenomenal quality is grasped as the quality that it is, that the language being used here will become intelligible. Finally, if and when you succeed in so suspending your normal objective relational interpretation of colour, you will discover something that you already knew (implicitly), something that you cannot directly express. It almost appears as if you have discovered nothing, or at least nothing worth mentioning. But, in *reality*, this is a pivotal event: you have discovered a phenomenon that lies outside the boundaries of physicality.

From here on we are speaking of the demonstration of pure phenomenal quality from *within* a phenomenological epochē, i.e. having already suspended the normal intentionality that only sees the coloured object that the quality of the colour presents. The task is to find a way to bring knowledge of this demonstration to explicit consciousness. Such “finding a way” is not prefigured, it is an *agreement* that must be reached before we attempt to speak of the experience. This agreement *is* the enacting of a phenomenological epochē. Given such an epochē, I can say that I now ‘know’ the quality ‘directly’, and not in the sense that I know an ordinary objectivity by means of an objective relational concept. There is no objectification that puts the quality ‘over there’ or as standing-against a subject who perceives or conceives the quality. It is rather that the phenomenal quality is known as the very means whereby an objective intentionality can become an object of conscious experience in the first place.

As Heidegger said (in a different context), we have ‘let that which shows itself be seen from itself in the very way in which it shows itself from itself’ (Heidegger 1962/2008, p. 58). There are no ‘hidden sides.’ The quality as quality is completely revealed in a way that cannot be doubted. Doubt can only arise in relation to a concept that refers outside itself.⁸ Whereas, my knowledge of pure phenomenal quality does not refer to anything outside the pure phenomenal quality itself - no ‘thing’ is the bearer of this quality and it does not exist as something separate from my knowledge of it.

This is not a question of correctly naming something. The pure quality has no name. To think that way is to confuse the quality with the objective relational colour instantiated in physical world. What fixes the reference of the colour in language are the physical conditions (my brain, the photons). We find no reason why we experience *this* quality rather than another, because all reasons are finally referred to the physical world and the physical world leaves the quality of the colour undetermined. From such a point of view *it doesn’t matter* what quality I experience, because whether my blue is your green makes no difference to our use of language.

However, the fact is that I ‘know’ what blueness is. This knowing is a direct intuition. Unless it is experienced as such, one has remained outside the epochē, one is *thinking* of the quality of blue using an objectifying concept that is a mere *sound*. Such thinking carries on regardless. Despite allowing there is such a thing as ‘what it is like’ to experience a colour quality, the experience itself is passed over, fixed in a certain framework of understanding that makes the quality itself inaccessible. Ordinary

⁸ Chalmers makes a similar point in his account of acquaintance (Chalmers 2010*b*, pp. 285-294).

language achieves this covering over because it fixes experience in objective relational terms. There is no ‘room’ for phenomenal quality, because such language refers to intra-worldly entities, and phenomenal colour is *not* an intra-worldly entity. This is how phenomenal quality *escapes* ordinary language while at the same time remaining an essential moment within the experience of using language to refer to colour, in that our experience of phenomenal colour *expresses* what it is we mean when we refer to colour. It is *that-by-means-of-which colour is made known to us as colour*.

To think that this direct ‘seeing’ of phenomenal colour can become a premise within a philosophical argument is to again misunderstand the demonstration. Philosophical argument is public argument concerning the determination of objective relational concepts. One can never demonstrate the essence of colour (‘prove’ that it ‘exists’) via such a procedure. To think so is to have failed to make the demonstration explicit and to deliberately attempt to understand the demonstration in such terms is to make the direct seeing, the essence of phenomenal quality, disappear. One only need consider the case of Frank Jackson, and the literature on the ‘knowledge argument’ to ‘see’ where such an approach leads.⁹

What is ‘seen’ in the demonstration, itself reveals that we can only speak of pure phenomenal colour indirectly, i.e. via metaphor and via negation. We have to “circle round” the phenomenon and catch it “out of the corner of the eye,” and then we have to say what it is *not*, i.e. it is not physical or intra-worldly or specifiable within objective relational concepts. This metaphorical circling and negation is a necessary consequence of the structure and limitations of ordinary objective relational language reference. However, to someone remaining within the stream of objective relational conceptualisation, it will appear that our inability to fix phenomenal quality in an adequate concept is a sign that we have not yet understood what it is, that we are dealing with some kind of fiction, or a phantom of the imagination. It is easy to become lost in such a line of thought. The answer is to stop and look again. The fact is there is direct intuition of phenomenal quality. It is true that phenomenal quality does not exist in the normal objective relational sense of the term. But phenomenal quality does ‘exist’ in the sense of being ‘knowable.’ Its existence *consists* in its being known. This existence is not something put together in the imagination. If phenomenal colour were something imagined then we should be free to vary the essence of colour in such a way as to imagine a new *kind* of colour, such as Terry Pratchett’s octarine.¹⁰ But this is not the case. The range of our phenomenal colour experience is strictly determined by the capacity of our nervous system to discriminate objective relational colour. We can, with Hume, imagine a *shade* of colour we have never actually seen, but we cannot imagine another *category* of colour.

⁹ I am referring here to Jackson’s intuition that colour experience involves a direct knowledge (Jackson 1982) and his subsequent reconceptualisation of that intuition as an illusion (Jackson 2003).

¹⁰ See (Pratchett 1985).

Finally, we must draw a distinction between my *being* conscious of a pure phenomenal quality, and my *seeing* that I am conscious of a pure phenomenal quality. These are the two crucial moments within the unity of a conscious experience of phenomenal quality. The first moment, my being conscious of the quality, is the experience that Chalmers is indicating in his account of pure phenomenal concepts. This experience is a *pre-reflective* registering of the (non-physical) quality that is the *content* of a pure phenomenal concept. In Chalmers' account, as content, the pure phenomenal quality is experienced but has no independent causative agency. In this way the phenomenal character of consciousness is acknowledged without violating the principle of causal closure.

However, Chalmers misses the second moment of consciousness. In order to deploy a pure phenomenal concept I must already have direct intuition of pure phenomenal quality, i.e. I must be able to consciously demonstrate such a quality – because the quality only ‘exists’ as a quality insofar as I experience it. It is this demonstration that enables me to ‘see’ the distinction upon which my concept of pure phenomenal quality is founded.

Without such a ‘seeing’ demonstration I will have no conception that there is any such ‘thing’ as a pure phenomenal quality, even though I passively (pre-reflectively) experience pure phenomenal quality as the content of pre-existing natural language concepts. It is here that physical causation enters the account. For, my ‘seeing’ of pure phenomenal quality depends on something that is not physical, i.e. pure phenomenal quality itself. This quality, as a phenomenal quality, according to causal closure, cannot be the cause of any physical event. And yet, my consciousness of this quality, i.e. my direct experience of pure phenomenal quality, *is* the cause of a physical event. That physical event is my forming a new conceptual category that recognises the existence of pure phenomenal quality on the basis of my consciously experiencing that quality. In order for this to happen, I not only have to pre-reflectively experience a pure phenomenal quality, I have to *see* that I am having such an experience. It is this *seeing* that is crucial for the negation of causal closure: it is where the direct intuition of phenomenal quality crosses over from being a passive (implicit) experiencing to being an explicit knowledge that makes a difference in the physical world. The essential point is that what is brought to explicit consciousness cannot itself be identified with the action of a physical process, i.e. a physical process cannot ‘see’ or represent an experience of phenomenal quality *as* an experience of phenomenal quality, it can only represent phenomenal quality as a physically instantiated objectivity. In contrast, my direct ‘seeing’ of phenomenal quality is a self-reflection of consciousness upon a pure experience of being conscious. As such it is nothing physical. And yet it makes a physical difference to my subsequent behaviour. I can now conceptually distinguish the being of pure phenomenal quality. Therefore the causal closure of the physical is false.

It should be noted that this seeing of pure phenomenal quality is not something that can be accounted for in terms of existing higher-order thought or higher-order

perception theories. Such theories assume that higher order brain processes are *caused* by preceding physical events occurring in the brain. Whereas my conscious seeing of a pure phenomenal quality is not a physical event. A higher order thought that is physically determined cannot think about a pure act of conscious seeing, because, as far as the physics is concerned, no such act has occurred.

7 The Validity of the Phenomenological Demonstration

Chalmers' account of phenomenal judgment is that the pure phenomenal content of a pure phenomenal concept does not cause our judgment that our colour experience has *this* pure phenomenal content. Our judgments are determined by the physics of the brain, and it just turns out (perhaps on the basis of psycho-physical law) that conscious entities, such as ourselves, have the experience of pure phenomenal content whenever we attend (correctly) to our immediate phenomenal experience. The contradiction in Chalmers' argument is that *if* the pure phenomenal content of a pure phenomenal concept is not able to independently cause any event in the brain, then there is no mechanism whereby we can be caused to *see* that there is any such content. Whereas we do see that there is such content, and our seeing has physical effects, viz. our thinking and speaking of our seeing.

Nevertheless, it is conceivable, on the basis of mimicry, that an unconscious entity could form a structure corresponding to my thought that "I am seeing the pure phenomenal quality of this colour now." Such an entity could then produce phrases of the same form in just those situations where I assert the same thing. In fact, we do not need to imagine unconscious entities to illustrate this point. It is quite conceivable in everyday life that there are people who will assert that they are seeing the pure phenomenal quality of a colour when in fact they are looking at a coloured object and repeating the phrase out of habit and without consciously appropriating the experience, i.e. without suspending their normal conceptualising of experience so that the quality of the colour is 'seen' as something distinct in itself.

In the case of someone who has *never* demonstrated the pure phenomenal quality of a colour to themselves, but speaks of such experience, their speech will be meaningless. It is not enough, as it is in the case of everyday reference, to defer to the knowledge of others. Despite my being immediately acquainted with colour throughout my conscious life, unless I have distinguished the pure phenomenal quality as something non-physical, then my reference will remain attached to the objectively physical entity that I take to be causing my experience. That entity may be the coloured thing in front of me, or the particular brain state that I believe is causing my experience. I can attempt, by close questioning, to try and elicit whether someone has really grasped what it means to see a pure phenomenal quality, but it is still possible that they are responding according to what other people have said on the matter. Until we have access to a brain-o-meter

that can detect a physical difference, we must acknowledge that the only person I can be certain has grasped this matter is *myself*.

To the person who has never consciously appropriated an experience of pure phenomenal quality, it will also appear quite conceivable that there is no such quality, and that all such talk has developed on the basis of mimicry and false belief. It is no good arguing that we could not have originally developed a concept of pure phenomenal quality unless there were such qualities. I can imagine many things that do not have any reality. The *only* ground for the assertion that there is more to our phenomenal experience than can be explained in objectively physical terms, is the immediate seeing of that experience. The essentially private and subjective nature of phenomenological experience means this cannot be demonstrated through a procedure of adversarial argument. There is nothing objective that can be pointed to as a standard of truth. Consequently, all disputes in this area finally resolve down to the question of whether the immediate evidence of a phenomenological demonstration is decisive.

8 The Phenomenological Consequences

To summarise, if we accept the division of reality into the objectively physical and the subjectively phenomenal, and also that the realm of the objectively physical is causally closed, then we have arrived at a contradiction. The contradiction concerns my conviction that there is more to phenomenal experience than can be explained in objectively physical terms. If causal closure were true, and the activities of my brain and body could not in any way be independently influenced by the phenomenal quality of my phenomenal experiences, it follows that I should be unable to form any notion of my having pure phenomenal experiences that are not completely identified with objectively physical states of the world. In that scenario, we would all agree with Dennett, and eliminative physicalism would be an obvious truth. Whereas the empirical evidence is that eliminative physicalism is not an obvious truth, i.e. *many* people have the conviction that there is more to experience than can be explained in objectively physical terms. If we accept the evidence of experience and thereby accept this conviction as true, its truth can only be justified on the basis of a direct knowledge of phenomenal experience. If there is direct knowledge of phenomenal experience, and I am able to state that knowledge, as I am doing now, it follows that my phenomenal experience, properly (consciously) demonstrated, has had an effect on my physical behaviour. That effect is direct and simple: I *see* that there is something more to my experience than can be explained in objectively physical terms. It is that very seeing that is the non-objectively-physical cause of certain objectively physical behaviours, viz. my forming the proposition that “there is more to experience than can be explained in physical terms” and my uttering of that proposition. If phenomenal experience can determine physical behaviour in this way, then the principle of causal closure is false.

The consequence of this result is to put the ontological foundations of objective physicalism into question. Epistemically, the division of reality into the objectively physical and the subjectively phenomenal has been extraordinarily successful in terms of the progress of the physical sciences. However, that success does not imply that reality is ontologically distinguished along the lines that physical science tacitly assumes. Our failure to give a coherent account of human consciousness in terms of the objective physicalist program is a clear indication that we may have the ontology wrong. If we relinquish the principle of causal closure and allow that my being conscious (in certain circumstances) has an influence on my behaviour, then it is no longer possible to clearly divide reality along the axis of objective physicalism. That axis distinguishes phenomenal experience on the basis of its causal dependence on the presence of objectively physical processes. It is this dependence that gives the objectively physical its ontological primacy. In denying causal closure, we also deny the coherence of attempting to explain phenomenal experience in entirely objectively physical terms. Phenomenologically speaking, objective physicalism is an unverifiable hypothesis concerning the possible structure of reality, a hypothesis that is negated on the basis of my phenomenological seeing that there is more to experience than can be explained in objectively physical terms.

References

- Chalmers, D. J. (1996), *The Conscious Mind*, Oxford University Press, Oxford.
- Chalmers, D. J. (2010a), The content of phenomenal concepts, in D. J. Chalmers, ed., 'The Character of Consciousness', Oxford University Press, Oxford, pp. 251–275.
- Chalmers, D. J. (2010b), The epistemology of phenomenal belief, in D. J. Chalmers, ed., 'The Character of Consciousness', Oxford University Press, Oxford, pp. 277–304.
- Dennett, D. C. (1991), *Consciousness explained*, Little, Brown and Company, Boston, Mass.
- Elitzur, A. C. (2009), Consciousness makes a difference: A reluctant dualist's confession, in A. Batthyány & A. C. Elitzur, eds, 'Irreducibly Conscious: Selected Papers on Consciousness', Universitätsverlag Winter, Heidelberg, pp. 43–72.
- Heidegger, M. (1962/2008), *Being and Time*, Harper Perennial, New York.
- Husserl, E. (1970/1992), *The Crisis of European Sciences and Transcendental Phenomenology*, Northwestern University Press, Evanston, Illinois.
- Jackson, F. (1982), 'Epiphenomenal qualia', *The Philosophical Quarterly* **32**(127), 127–136.
- Jackson, F. (2003), Mind and illusion, in A. O'Hear, ed., 'Minds and Persons', Cambridge University Press, Cambridge, pp. 251–272.
- Montero, B. & Papineau, D. (2005), 'A defence of the *via negativa* argument for physicalism', *Analysis* **65**(3), 233–237.
- Popper, K. (1959/2002), *The Logic of Scientific Discovery*, Routledge Classics, New York.
- Pratchett, T. (1985), *The Colour of Magic*, Corgi Books, London.
- Strawson, G. (2008), Real materialism, in G. Strawson, ed., 'Real Materialism and Other Essays', Clarendon Press, Oxford, pp. 19–52.
- Wittgenstein, L. (2001), *Philosophical Investigations*, Blackwell, Oxford.