

Kim Davies

Emergence from What?

A Transcendental Understanding of the Place of Consciousness

Abstract: *This paper argues that the standard formulations of the question of how consciousness emerges, both synchronically and diachronically, from the physical world necessarily use a concept of the physical without either a clear grasp of the concept or an understanding of the necessary conditions of its possibility. This concept will be elucidated and some of the necessary conditions of its possibility explored, clarifying the place of the mental and the physical as abstractions from the totality of an agent engaged in the life-world. The notion of a disruption or breakdown in the agent's normal engagements in the world will play a key role in the argument, which in turn provides a transcendental underpinning to recent enactive and embodied theories of mind by exploring some of the necessary conditions of being an agent in the world.*

1. Introduction

There is at least some agreement over the *prima facie* existence of a question (or closely related set of questions), normally referred to as the 'hard problem of consciousness' or the 'explanatory gap', namely to understand how consciousness 'arises from' our physical bodies.¹

Correspondence:
Email: kimdavies1951@gmail.com

[1] Standard formulations include the following: 'How can living physical bodies in the physical world produce such phenomena [of consciousness]?' (Dennett, 1991, p. 25); 'How and why consciousness arises from physical processes in the brain?' (Chalmers, 2010, p. 399); '[We lack] an explanation for how our physical constitution gives rise to consciousness' (Levine, 2001, p. 78); 'How can technicolor phenomenology arise from soggy gray matter?' (McGinn, 1991, p. 349).

In this paper I shall label this the Question of the Emergence of Consciousness (hereafter QEC), so that it includes both the synchronic and the diachronic or historical emergence of consciousness; the use of the term ‘emergence’ is informal and not tied to any of the various theories of epistemological or ontological emergence. There is also a general acceptance that neurophysiology/phenomenology is advancing our understanding of the correlations between processes and events in the body (including the brain) and processes and events in the mind. Thereafter views diverge with respect to whether this is a hard, (relatively) easy, or impossible question; on the foundations of the problem; and on the various approaches to dealing with it. Common ground amongst most disputants is a level of ontological security accorded to the physical realm, without, I will argue, a clear grasp of either the concept of the physical required to frame the QEC, or of the necessary conditions of the possibility of that concept.² This holds most obviously for physicalists, but also for dualists and for panpsychists who attribute mentalistic properties to fundamental physical entities. Excepted in particular from this consensus (along with radical idealists and some monists) are those phenomenologists who see the QEC as arising from the unreflective adoption of an objective stance which ignores the nature of experience which makes such a stance possible: their position on the QEC will be briefly considered in Section 6.

Transcendental arguments can be understood as ‘indispensability chains’ (Taylor, 1978–79): they begin with a starting point assumed by a theory and go on to show how the necessary conditions of that starting point undermine the theory itself. Although their use is most familiar in arguments against forms of scepticism,³ they can also be used against other theses to indicate their incoherence and, as here, to advance alternatives. Any transcendental argument can be interrogated as to its initial starting point and to each specific link in the proposed chain of necessary conditions: the aim in this paper is to lay out a direction of argument that is robust and interesting enough to prompt further discussion.

The paper begins by elucidating the concept or understanding of the physical presupposed by the standard formulations of the QEC. From this starting point it will then analyse some of the necessary conditions

[2] See, for example, Chalmers (2010, p. 104) where he divides the most important views on the metaphysics of consciousness into those which require no expansion of a physical ontology, and those which require expansion or reconception of a physical ontology.

[3] With debatable success: see Stern (1999) for a useful survey, and Brueckner (2010, Part 1) for sustained discussion.

of the possibility of the understanding of the physical, outlining a transcendental argument to clarify the place of the mental and the physical as abstractions from the totality of an agent engaged in the life-world, with the notion of a disruption or breakdown in that engagement playing a key role. By exploring some of the necessary conditions of being an agent in the world, this in turn provides a transcendental underpinning to recent enactive and embodied theories of mind.

2. Understanding the Physical: A Review

The issue of what we mean by the term ‘physical’, or what concept is articulated through that term, or what it is to be physical, has received some attention over the last thirty years or so but is still relatively marginal to the philosophy of mind: most philosophers in the field take for granted that we know what we mean and move on. But in taking the concept for granted in this way a number of presuppositions are necessarily made which determine the framing of the QEC. Our understanding of the physical is central not only to physicalism but to those approaches which to some extent define themselves in opposition to physicalism. After all, we should, as Montero (2001, p. 62) notes, ‘have at least a rough idea of what it means to be physical, not necessarily a strict definition, but at least a notion of the physical that excludes some, if not actual, then at least possible, phenomena from being physical. For if we cannot even conceive of something being nonphysical, it is difficult to grasp what physicalists could be arguing for — to say nothing of what they could be arguing against’. Consider then some recent approaches to our conception of the physical. Amongst the candidates we can distinguish positive, negative, and combinatorial approaches. The predominant positive conceptions are as follows:

Theory-based Conceptions: A property/entity is physical if and only if it is the kind of property/entity that physical theory/physics tells us about, or metaphysically (or logically) supervenes on such properties/entities.

Object-based Conceptions: A property is physical if and only if it is either the sort of property required by a complete account of the intrinsic nature of paradigmatic physical objects and their constituents, or is a property which metaphysically (or logically) supervenes on such a property.⁴

[4] These formulations are from Stoljar (2009).

Now the main focus of discussion around these conceptions has so far concerned Hempel's Dilemma and how to avoid it,⁵ with the problem of circularity receiving relatively short shrift (e.g. Stoljar, 2009, Section 11.1). Nevertheless it is circularity which I shall argue is the fundamental problem here. It is important to stress that the issue here is not the failure to provide a non-circular *definition* of the term 'physical' — there are many words that we come to understand without such definitions. Nor is it the failure to provide a non-circular reductive analysis of the concept — there are many concepts we understand without being able to provide such analyses. The problem is that the approaches in the literature can give us an understanding of the notion of the physical only if such an understanding is already presupposed, and thus they fail to explicate this understanding at all.

With the Theory-based Conception the issue arises when we ask for an understanding of the 'physical theory' or 'physics' which figures centrally in the account of what it means to be physical and which looks to introduce an obvious circularity. For even if we aim at providing understanding, rather than strict definition, the theory-based conception requires a pre-understanding of physical theory/physics. Either this will itself already presuppose an understanding of the physical, and so introduce circularity, or will rely on paradigmatic examples (e.g. Newtonian, Einsteinian theories) perhaps in association with the notion of 'family resemblance' popularized by Wittgenstein.⁶ However, such an approach will only work if we at least implicitly understand why other theories with which the physicalist is likely to be uncomfortable (e.g. Aristotle's Physics, Whitehead's process theory) are ruled out as paradigmatic physical theories,⁷ and barring a hapless nominalism regarding what is studied in university 'physics' departments this will necessarily be based on a

-
- [5] See Hempel (1980). Briefly, that dilemma holds that a) if we define the physical to be what current physics says there is, then physicalism is very likely false, and b) if we define the physical to be what an ideal/complete physics says there is then physicalism lacks determinate content (and if the ideal physics were to include mental entities or properties then physicalism would be uninterestingly trivial). Strictly speaking this concerns the Theory-based Conception, but a close parallel can be devised for the Object-based Conception since paradigmatically 'physical' objects may turn out to have mental properties also.
- [6] It could be argued that understanding a concept is arguably not an either/or affair, and the appeal to physics or physical theory is merely an initial route to clarifying a pre-scientific notion. Even so, to avoid circularity, the appeal must ultimately draw on the pre-theoretic understanding of the physical, with the consequences discussed in Section 3 below.
- [7] See Crane and Mellor (1990) for a discussion of the attempt to distinguish between psychology and the physical sciences. It could perhaps be argued that neither Aristotle nor Whitehead produced *scientific* theories, given the characteristic features of such theories (e.g. Dowell, 2006, Section III), but this is to say the least arguable and in any case fails to

pre-understanding of the physical which is yet to be explicated.⁸ A further move here is to attempt to define the physical in terms of the hallmarks of physics or physical theories (e.g. Poland, 1994, Chapter 3; Dowell, 2006). Now it is far from clear that this would rule out Aristotle or Whitehead's teleological theories, and in any case there is a deeper issue. For example, Dowell's proposal, which builds on Poland's, defines "the complete and ideal physical theory" as the complete and ideal scientific theory of the world's most fundamental elements' and argues that 'to count as basic and physical, a property must be well-integrated into the most complete and unified explanation possible for the relatively most basic occupants of space-time'. However, although Dowell admits to not saying anything much about what makes an element 'fundamental', in the context it seems clear that she has already excluded, for example, the most fundamental elements of my phenomenal consciousness, or the most basic qualia that constitute my current experience from the category of the world's most fundamental elements.⁹ The notions of fundamentality and basicness are here already linked to the notion of physical things, and so cannot serve to explicate the concept of the physical.

A parallel circularity problem arises with the Object-based Conception. On what basis are, say, rocks or tables or electrons selected as paradigmatic physical objects and others, e.g. hallucinations, people, colour sensations, not? The acceptance of paradigmatic physical objects is not merely a decision to agree with a certain selection: other objects can be ruled out as paradigms only on the basis of a shared pre-understanding of the physical of which the paradigmatic objects are clear examples. In ordinary life it is a relatively straightforward matter to explicate this pre-understanding: physical objects are those things such as apples and rocks that we can see, touch, and manipulate, and which can impact on us to good or ill effect — apples sustain us, rocks can injure us. We can then move on from the paradigm examples to discuss how to account for non-paradigmatic objects, e.g. emails, or rainbows. But this ordinary understanding is normally ignored in the literature, or dismissed as crude, pre-scientific 'folk wisdom' which, even if a necessary prerequisite to science, needs to

rule out the possibility of scientific theories which are either teleological or incorporate fundamental mentality.

[8] See Dowell (2006) for further discussion of this haplessness.

[9] For example, she sees it as an advantage of her formulation that 'its ontological commitments are tied to the posits of theories that are confirmed in accordance with our best methods for justifying our beliefs about the natural world'.

make way for the theoretical clarity provided by physical science. We are still no clearer to a non-circular understanding of the physical.

If a positive understanding of the physical via paradigmatic theories or objects runs into the problem of circularity, in that it already presupposes an understanding of the physical which is not made clear, then perhaps we can grasp the physical in its opposition to other notions. In particular, can the physical be understood as that which is non-mental? This is the *via negativa* approach taken by Montero and Papineau (2005), which essentially takes the causal argument for physicalism and replaces 'physical' with 'fundamentally non-mental' giving us a version of physicalism summarized in the proposition that the mental is, at its most fundamental level, entirely non-mental.

There are objections in the literature regarding the effectiveness of this approach in evading Hempel's Dilemma (e.g. Dowell, 2006; Judisch, 2008) and it seems to have the discomfiting consequence of according the mental at least an epistemological and semantic priority over the physical. Nevertheless the main focus here is again on circularity. For the approach to be successful, it assumes that we have a grasp of the mental which is independent of our grasp of the physical, and this seems optimistic. Not only do dictionary definitions of 'mental' and 'mind' cast them in contradistinction to 'body' and 'matter', but arguments such as the Private Language Argument from Wittgenstein (1953) should at least make us wary of assuming that it is possible to have an understanding of mental events independently of our grasp of our intersubjective engagements with the world. It is hard to see how the usual understandings of the mental in terms of consciousness, intentionality, sentience, experience, and so forth are possible without explicitly separating them out from the bodily aspects of our being. Nevertheless, a comprehensive argument to show that an understanding of the mental presupposes a grasp of the physical is beyond the current paper, although more will be said on the subject in Section 6. For now, the indication that the circularity of understanding on which I have so far focused cannot be avoided by the *via negativa* will need to suffice.

There exist in the literature attempts to combine positive and negative approaches such as that from Wilson (2006), who combines an appeal to fundamental physics with a 'No Fundamental Mentality' constraint. While this may have merits as a constructive way of addressing Hempel's Dilemma, it runs into the circularity issue on both sides as it presupposes an understanding of both the physical and the mental, and cannot help with the key issue of explaining the understanding of the physical required by the QEC.

Following this brief overview of current approaches at least three initial responses may arise. Firstly, neuroscientists and neurophenomenologists may express their lack of interest in the circularity issue: they are interested only in the physical processes that correlate to mental phenomena. This is fine: it is valuable work and generates in itself no philosophical problems and merely sets the problem aside.¹⁰ Secondly, philosophers may declare their frustration with all this talk of circles of understanding by saying that we clearly break into them somehow and so should put them aside and move on. This attempted dismissal will not do: the necessary conditions of breaking into this circularity of understanding have important philosophical consequences which this paper will attempt to draw out. Thirdly, it may seem that there is no way of determining which understanding of the concept ‘physical’ is the correct one so that we are stuck in a sterile terminological debate with no clear way forward (Chalmers, 2011). The response here is to identify the role which the concept of the physical is required to play in the various discussions of the QEC and use that as the guide to clarification of the concept. It is to this identification and clarification we now turn.

3. Understanding the Physical: Pre-theoretical Requirements

The attempts outlined above to explain the concept of the physical used in framing the QEC all require a pre-theoretical understanding of ‘physical’ and/or ‘mental’. Our everyday understandings of what we mean by the physical are too loose and indeterminate to give us a clear lead here, although the ‘folk wisdom’ mentioned in the previous section will have a part to play. The place to start is by clarifying the understanding of the physical that is required if discussions of the QEC are to get going in the first place. For current philosophers much of the underlying attraction of physicalism, and so the importance of the notion of the physical in framing the QEC, is provided by the continuing development of the physical sciences, with physics pre-eminent in the pantheon (e.g. Papineau, 2001). Modern physics explores the fundamental components of the physical world, their structures and causal relationships, and aims to show us how things work from the microphysical level, to the level of medium-sized dry goods, to the inter-galactic level. It has gained its credibility through increasingly

[10] Although the basis for the neurophenomenological approach is more sophisticated than here presented. See Chalmers (1997), Varela (1997), and Thompson (2007, Chapter 11) for examples.

sophisticated experimentation, observation, theory-construction, prediction, and practical utility. Broadly speaking, physical theory over the last five centuries postulates a range of entities, a set of quantifiable properties (e.g. energy, momentum, charge), and a set of forces (e.g. gravitational, electromagnetic, strong and weak interactive) which mediate the distribution and exchange of these quantifiable properties. While the specifics of the theories may change, this précis provides for the continuity in physical theory that both underpins the physicalist position and helps explain the intuition that our understanding of the term ‘physical’ is linked to whatever it is our current ‘well-confirmed physical theory tells us there is’ (Dowell, 2006).¹¹ It does not, however, avoid the issue of circularity articulated earlier (it takes for granted our understanding of ‘physics’ and ‘physical world’), and to do this we must attempt to explicate the necessary pre-theoretic understandings on which physical theory depends.

To begin with, it is clear that for physical theory to exist and develop, entities in the world must be pre-understood as in principle (directly or indirectly) manipulable, and publicly observable, otherwise there could be no observation, experimentation, or objectivity in science: the ‘folk wisdom’ described earlier has its place as a necessary prerequisite of physical theory. However, in addressing these things with which we engage in the world, physics then takes up a particular stance which constitutes a more specific pre-theoretical understanding. It addresses things only in respect of those properties which exist independently of the subjects which come to know them and which can be objectively identified and quantified, such as mass, location, extension. It does not concern itself with those properties whose existence depends essentially on subjectivity (or intersubjectivity) and which cannot be objectively identified and measured, such as meanings, values, purposes, and sense-qualities (thereby excluding Locke’s ‘secondary qualities’). This focus on the objective properties of things in the world is not an accidental feature of physics but lies at the core of its methodology and its credibility. It is crucial to the intersubjective verifiability at the core of the ‘third-person’ outlook which scientific theory takes for granted (*cf.* Dennett, 1991, pp. 66–71). If it were suggested that certain events could be explained only by postulating that a certain object or event had a certain meaning or purpose, the physicist would count this as a mark of an area of incompleteness in physics, not as an addition to its theoretical resources. Meanings,

[11] This then motivates Hempel’s Dilemma, for how can we know that a future/complete physics will not postulate extensions to the theory such as proto-qualia or micro-volitions which any physicalist worth the name will deem inappropriate?

values, purposes, and sense-qualities are the objects of other fields of study but not of physical theory.

For closely related reasons physical science itself¹² makes no attempt to investigate mental phenomena as such (rather than their physical correlates) since they are neither publicly observable in the ways required by physical science nor appropriately located in physical space (*cf.* McGinn, 1997, p. 99, who explains our pre-theoretical location of conscious events roughly in the vicinity of the brain as a courtesy location, parasitic on the location of the body: ‘there is no independent route on to mental location’).¹³ This means that this exclusion of the mental is *a priori* as far as physics is concerned (ruling out the spectre of ‘inappropriate extensions’ to physics raised by Hempel’s Dilemma).¹⁴ Any instances of mental causation as such will therefore be invisible to physical science and be accorded no place in its theories (*cf.* Lowe, 2008, p. 74).¹⁵ This double exclusion of mental properties and entities from the domain of physical science gives us an initial grasp of the oppositional relationship between the mental and the physical which the *via negativa* hoped to exploit, and must lie albeit implicitly at the core of the understanding of the physical on which physical theory and, by extension, the QEC is based. Without such an understanding no sense could be made of the projects of the physical sciences, and no credence given to their results. Physicalism itself would be unable to draw on science as its primary rationale and would have no obvious attractions. It is hardly surprising then that the QEC seems so hard: it is an attempt to bridge a gap between two aspects of the world the concepts of which are *a priori* mutually exclusive. It is with this double-faceted understanding then that we can develop a transcendental understanding of the place of consciousness.

[12] The emphasis on physical science *itself* is intended to distinguish the science from any metaphysical claims or assumptions about identity, realization, emergence, or supervenience which scientists themselves may wittingly or unwittingly make.

[13] This is not to deny that consciousness is in some sense publicly accessible, through the third-person reporting that is the basis of Dennett’s notion of ‘heterophenomenology’ in Dennett (1991, Chapter 4).

[14] See, for example, Dowell (2006) for a defence of the view that the exclusion is contingent and *a posteriori*.

[15] This also holds for spirits, angels, gods, etc. since whatever their ontological status, their actions can only be understood in terms of a form of mental causation (*cf.* Worley, 2006).

4. Necessary Conditions of the Possibility of Understanding the Physical

The conclusion from the discussion in Section 2 was that neither a positive understanding of the physical (via ostensive examples of paradigmatic physical theories or objects) nor a negative understanding (as non-mental) can succeed. Positive approaches already presuppose an understanding of the physical, while negative approaches presuppose a prior understanding of other categories (e.g. the mental) which is at least as problematic, and which in any case also requires a prior understanding of the physical. In Section 3 the necessary pre-theoretical understanding required for the possibility of physical theory was outlined. Let us now consider the necessary conditions of the possibility of this pre-theoretic understanding of the entities in the world studied by physical science as manipulable, and publicly observable, with the requirement that only the subject-independent properties of those entities be directly addressed.

If we consider the required minimal pre-understanding of entities subject to investigation by the physical sciences as those which can be manipulated and observed by an agent, and which are not subject-dependent, the question still arises as to how that understanding is possible in a way that enables the business of physical science to get going.

For our primary understanding of things in the world and their properties cannot be an understanding of them *as* the objects and properties described by physical science, as this would place us in the inextricable circularity discussed in Section 2 above. The processes of physical science itself require a pre-theoretic understanding of the entities which are manipulated, observed, and thought about by, amongst others, scientists. Furthermore, on one understanding of Wittgenstein's (1953) argument against the possibility of a private language, the use of scientific concepts presupposes a prior communication community of such agents engaged in 'language games' — embedded in 'forms of life' — through which the scientific terms themselves acquire meaning. These forms of life require a primary understanding of entities in the world which is related to the shared purposes of the agents which form the relevant community.

While a full discussion of this view is beyond the scope of this paper,¹⁶ this primary understanding of things in the world must be of them as having a role related to the pre-theoretic purposes in the light

[16] A fuller exposition of this line of argument is provided in Apel (1980, especially Chapter 5). The view also has strong affinities with Husserl's concept of *Lebenswelt* or 'life-

of which the agent acts: berries, for example, are understood as a kind of food, stones as potential weapons, a flooded river as an obstacle to progress, etc.¹⁷ But if the primary understanding of things in the world is in relation to our pre-theoretic ends and purposes then how is the understanding of them specifically as *physical* entities, that is as things with objectively measurable properties and stripped of subject-dependent qualities, possible?

If, as argued in Section 2, neither positive approaches nor the *via negativa* can succeed as a way of explaining our understanding of the physical, then we are in a situation from which there is only one exit route. The primordial understanding of the physical is possible only through reflection on a breakdown in a pre-existing totality from which the notion of the physical as such can be abstracted. This is a situation of breakdown in the normal processes through which agents pursue their purposes in their environment, and in relation to which they understand things in that environment. Thus the stick we are using breaks, the crops fail, the water is polluted, etc. Such breakdowns bring to our attention our expectations (implicit or explicit) concerning things in the world. They also allow us to focus on those things with a view to finding out what went wrong and fixing it: that is, with an interest in 'technical control' (Habermas, 1972). So we can abstract from, or overlook, the purpose, the beauty, the meaning, or the value of the thing in question and treat it purely in terms of its physical structure and properties.¹⁸ Without such disruptions in our normal engagements there is no way of arriving at an understanding of the physical as that aspect of things within the world from which all subject-dependent properties have been stripped. It is at this stage that science can begin to analyse the causal structures that underpin and make possible our ordinary dealings with the world, in the interest of practical control of our environment. This will include exploration of the physical structure of the agent itself: here we overlook the agent's purposes and abstract from the body as the mode of the agent's engagement with the world, so that we can focus on its physical aspect with a view to fixing it when things go wrong: for example when a bone is broken or the brain is damaged (an example addressed in more detail in Section 6). We can then extend our scientific explorations

world' as described in Husserl (1970), and I shall use his term as a shorthand where convenient.

[17] This line of thought emerges clearly in Heidegger (1962). It also underpins the notion of 'forms of life' in Wittgenstein (1953) and of 'affordances' in Gibson (1979).

[18] Heidegger uses similar examples in his discussion of the distinction between 'ready-to-hand' and 'present-at-hand' in Heidegger (1962).

beyond local practical interest to theories about the nature of the physical realm which refer to quarks and neurons, leptons and synapses (and this leads us to the theory-based notion of the physical which underpins Hempel's Dilemma). Nevertheless, however far scientific theories may take us from our everyday concerns, and however far they penetrate the composition of the material realm, they cannot justify the claim that the physical nature of things exhausts the world. For physical science necessarily presupposes the existence of a world of affordances,¹⁹ meanings, and uses within which purposeful agents are situated and act: it explores that set of causal properties, structures, and processes that we focus on when we abstract from those affordances, meanings, and uses. In so far as we are ontologically committed to the entities and properties described by the physical sciences we are also already ontologically committed to that richer meaningful world and its agents.

This point needs clarifying. First, it is not merely a statement of the point that our commitment to the existence of the richer life-world is required by any practice whatsoever of ontological commitment. Rather, our understanding of what it means to be physical is possible only through reflective abstraction from that richer world of meanings, purposes, and values whose existence is therefore necessarily presupposed, although not explored by physical science itself. Second, this process of abstraction reveals that the physical must be understood as one aspect only of the agent and its life-world. Third, despite this two-fold priority of the agent and life-world there is no dispute here that physical entities existed prior to the existence of the agents from whose world the notion of the physical is abstracted: that existential priority is accepted. The consequences of these points for our understanding of the historical emergence of consciousness will be discussed in Section 6. Fourth, it is worth noting the broad point that there are pairs of general categories which can only be understood to some extent in opposition to each other, which cannot be grasped in either the positive or purely negative manner described in Section 2, and can only be understood as interrelated and opposed elements of a disrupted totality from which they are reflectively abstracted. Furthermore, once the operation of abstraction from a prior totality is performed in the interest of analysis, it is impossible to successfully reconstitute the totality by merely combining the elements which have been so abstracted. In the current case we cannot understand the world

[19] Gibson (1979, p. 18) sees affordances as 'offerings of nature... possibilities or opportunities' for animal life.

we inhabit as a set of physical entities to which meanings, purposes, and values have somehow been stuck on.

The centrality of the agent is also made clear by reflection on mental events. We can describe their intentional, representational, and phenomenal character, but this all relates to their content. It is famously harder to reflect on consciousness itself:²⁰ we do not ‘see’, or become otherwise aware of, a semi-transparent container of the contents, or a kind of ray directed towards them. Consciousness is not ontologically self-sufficient: it can only be referred to as ontologically dependent on something else, as a modification or attribute of that something else. The usual term for this ‘something else’ is the subject of consciousness, and considerable philosophical work has been devoted to an understanding of this subject.²¹ Nevertheless, whatever theory of the subject we adhere to, the subject cannot primarily be grasped as something to which an act of consciousness or a mental event is just added. For this would require just that understanding of the mental of which we are seeking an account. The subject must be grasped in a mode in which the mental, or the act of consciousness, is already embodied in the subject’s acting, thinking, perceiving, hoping, willing, feeling, etc. A necessary condition of our primary understanding of the mental is therefore abstraction from this engaged subject. Although there is no space to explore the issue here, there are arguments to the effect that key mentalistic concepts such as intention, belief, and speaker meaning are possible only through reflection on a failure of the agent to act in accordance with social norms, so that the development of the rich notion of mind we currently possess itself requires reflection on disruption in the normal patterns of engagement of the agent in the social world.

Before considering how the argument of this paper offers an approach to the QEC, let us take stock.

5. A Summary and an Objection

We can summarize the argument so far in the following propositions;

[20] The notions of the mental and consciousness are variously construed. While I understand the mental broadly to include consciousness there are clearly mental properties which do not reach consciousness.

[21] The terms ‘subject’, ‘agent’, ‘person’, and ‘self’ are used more or less interchangeably in this paper, varying only where a term seems to connote features more relevant in the context.

- 1) The standard framing of the QEC requires a concept of the physical which links it to the successes of physics and the physical sciences.
- 2) Current attempts to explicate this concept are circular and require a pre-theoretic understanding of the physical to succeed.
- 3) Physical theory necessarily presupposes an understanding of things in the world as in principle observable, manipulable, and publicly accessible.
- 4) The pre-theoretic understanding of the physical required by physics, and by extension the framing of the QEC is such that things in the world are considered only with respect to their objectively identifiable and quantifiable properties: neither subject-dependent properties nor mental entities are objects of study or theoretical postulates in physics.
- 5) The QEC arises as an attempt to bridge a gap between two aspects of the world the concepts of which are *a priori* mutually exclusive.
- 6) The necessary conditions of the required pre-theoretic understanding of physics include reflection on situations where our normal engagement with the world is disrupted due to a failure of something in the world or agent to operate normally. The notion of the physical is then arrived at through abstraction from all the subject-dependent properties of the thing in question.
- 7) Physical science (and so physicalism) cannot give an adequate account of either things in the world or the agent because the necessary conditions of its possibility include:
 - i) The existence of a world of which the physical aspect (entities, events, properties) is only one aspect. The physical cannot exhaust the world.
 - ii) The existence of the agent whose purposes and mental properties cannot as such be explored by physical science.
- 8) A necessary condition of the primary understanding of the mental is the ascription of mental features to the subject from which they are then reflectively abstracted.

Each of these propositions is open to interrogation in a number of ways but I will focus initially on one key objection.

This objection suggests that the argument confuses epistemological requirements with metaphysical ones, conceptual commitments with ontological ones, for surely we cannot reach valid conclusions about the nature of the real world just through considerations of the necessary conditions of meaning what we want to mean. This point must be

resisted: whenever the concept of the physical is meaningfully employed in the way required to frame the QEC it presupposes an ontological commitment to a purposeful agent and life-world from which we abstract its purely physical properties.

It is important to stress that this is not merely a conclusion about how we must *think of* or *conceptualize* the world, allowing reality itself to be characterized as ultimately physical. Whenever we use the concept or term ‘physical’ we are committed to the existence of the agent and its world of affordances as a necessary condition of its content or meaning. To then repudiate that necessary condition is to render propositions using the term ‘physical’ meaningless, and therefore the standard framing of the QEC meaningless.

Thus far the force of the argument presses most clearly against physicalism; since it claims that we cannot accept that there is nothing ‘over and above’ the physical because the use of that concept presupposes the existence of the agent and life-world of which the physical is only one aspect. In order to challenge the argument, the physicalist needs to either provide an alternative characterization of the concept of the physical that physicalism requires, or an alternative explanation of how that concept is possible without the ontological commitments I have argued it necessarily presupposes.²² Otherwise physicalism will appear guilty of repudiating one of the necessary conditions of its own meaningfulness and thus courting incoherence.

A further move, however, begins by accepting those commitments, that our grasp of the physical is possible only through reflection on breakdowns in the normal operations of the pre-existent life-world, and that the physical is only one aspect of that world. It then moves on to claim that this acceptance that does not rule out the possibility that the physical is nevertheless ontologically fundamental in so far as it metaphysically necessitates, through brute supervenience (*cf.* McLaughlin and Bennett, 2011) or grounding (*cf.* Trogdon, 2013), the other aspects of the life-world and the existence of consciousness. After all, the existential priority of physical entities has already been noted. Even so, such theories can provide no explanation of how or why the physical world metaphysically necessitates consciousness, and so a version of the QEC will remain.²³ Let us then leave such theories and look at how the argument of this paper addresses the QEC.

[22] I assume here the failure of general scepticism about the existence of the ‘external world’ since this would just put the physicalist in an even worse position.

[23] For an argument to this effect see O’Connail (2014a).

6. Accounting for the Emergence of Consciousness

We must begin by rejecting the QEC as a legitimate question, at least in its standard formulations (see footnote 1), as its central concepts presuppose the existence of a totality of agent and life-world (Heidegger's, 1962, 'Being-in-the-World'). Consciousness does not 'arise from' the physical world but must be seen as an abstraction from the totality of Being-in-the-World of which the physical is itself but one aspect. Let us now consider how the exploration of the necessary conditions of the possibility of the understanding of the physical offers a more productive approach to the issues of both the synchronic and diachronic emergence of consciousness.

To begin with the question of synchronic emergence, even if the QEC is rejected the question of how we should understand the apparent dependence of the mind on the body (including the brain) remains. On the current account both consciousness and the physical are abstractions from the agent in its life-world, and it is to further consideration of this agent, whose body with its range of possibilities of movement, action, and perception is the means by which the agent engages purposefully with things in the world (*cf.* Noë, 2009, p. 77), that we turn to indicate the following line of approach.

We can identify the following features required for the existence of an agent purposefully engaged in its life-world:²⁴

- a) The agent must be self-sustaining with a basic interest in carrying on being (recognized by Spinoza, 1985, Part III in the concept of *conatus*, and Heidegger, 1962, H.84 in the concept of the 'for-the-sake-of-which').
- b) It must possess its identity through being marked off from its environment.
- c) It must be able to perceive and make sense of its environment in the light of what is good for its continuation.
- d) It must be able to sense when its well-being or equilibrium is being disrupted, and respond appropriately.

This bare characterization enables us to understand the necessary connection between mind and body that is so elusive.²⁵ Feature b) establishes the interior of the self in relation to its exterior; feature c) estab-

[24] It is beyond the scope of this paper to argue for these points in detail, but see the discussion in Thompson (2007, Chapter 6).

[25] *Cf.* Nagel (2002), where he rejects the conceptual reducibility of mental concepts to physical ones, but suggests that we should look for new concepts of a common basis of the mental and the physical in order to account for the apparently metaphysically necessary links between the two.

lishes the need for some form of cognition; and feature d) establishes the necessary ability to sense the internal state of affairs of the self. Therefore an essential characterization of the self or agent, which is necessarily presupposed by the use of the concept of the physical, is that it both has a presence in an external environment (and thus a spatio-temporal location) and has an interior life which is not publicly observable in the same manner. It is the necessary co-presence of this interiority and exteriority which grounds the possibility of abstracting the mental and physical aspects of the self. Without these features there could be no agent engaged in the life-world. The apparent dependence of mind on body can then be understood as the felt registration of disequilibrium (e.g. hunger) or the restoration of equilibrium (the satisfaction of a good meal), or as a registration of the fact that something is going wrong with body (e.g. the pain of a sprained ankle, the sickness following food poisoning). This should be seen as the self registering its situation with respect to its self-maintenance in its environment, rather than as the dependence of one abstracted aspect — the mind — on another — the body. When science treats the body purely as a physical object, following reflection on breakdown in the normal pattern of engagements with the world, it can explore the causal connections which are at work with an interest in fixing things which go wrong (splints for broken bones, antibiotics for food poisoning, etc.).

A particular example may help to illustrate this approach to the mind/body relationship. Research into Alzheimer's disease may identify a causal link between the growth of plaques and tangles in the brain and the death of brain cells. Neuroscience can explain how physical events in or out of the body cause other physical events in the body, and this is what is happening physically (and it can perhaps be remedied) when part of your body (including your brain) is failing you. This process necessarily presupposes a primary grasp of the situation in terms of breakdown in normal engagements (e.g. 'I can't find my way home', 'I can't recognize my family'). The science itself explains only how certain physical events cause other physical events: it cannot itself account for the correlation with the symptoms of Alzheimer's which are defined in normative terms of what is required for a functioning agent in its (social) world.

The way to conceive of the synchronic emergence of consciousness, then, is to see it not as an issue of how one abstracted feature of the agent arises from another abstracted feature, but as a clarification of the interrelation of the key features of the self outlined above in the

context of that being-in-the-world which is a necessary condition of the possibility of the concepts of the physical and the mental.

If these necessarily brief remarks are enough to indicate how an understanding of the apparent dependence of mind on body can be developed, what can be said of the obverse, the mental causation of physical events? Attempts to explain the apparent intervention of the mental in the supposedly closed world of physical causation have generated arguments with increasingly sophisticated technical apparatus (*cf.* papers in Gibb, Lowe and Ingthorsson, 2013) all of which take for granted the concept of the physical. The alternative picture described above provides room for the following account. The concept of the physical required by physicalism is possible only on condition of a prior commitment to a world of affordances with which an embodied subject actively engages, and of which the purely physical is revealed as only one aspect. We act in the world of our own volition in ways we can explain non-causally when explanation is called for (e.g. when our behaviour transgresses implicit or explicit norms) by reference to our beliefs and desires. There is no need or possibility here of claiming the existence of mental forces or events which cause physical events — there is no ‘mental causation’ in this sense — because the categories of mental and physical have been separated as ontologically distinct. Physical science describes and explains what is happening physically when we do so act, but presupposes the agent acting in the life-world of which the physical is only one aspect.

Moving on to the question of the diachronic emergence of consciousness: the question of the historical emergence of the mental. What is not in dispute here is that the world has a history which goes back beyond our adult life, our conception, human life, recognizable life forms, and the solar system to the beginnings of the universe. Neither can it be in dispute that the universe had the potential to develop conscious beings and the affordances they take advantage of.²⁶ The current account avoids the difficulties of the QEC by again taking its direction from a consideration of the agent in its life-world which is necessarily presupposed by the understanding of the physical.

This approach means that in order to understand the historical emergence of consciousness we must reverse the normal sequence of attempted explanation. Instead of beginning with purely physical forces and trying to work out how they give rise to consciousness we must begin with a clear understanding of ourselves as agents and work

[26] *Cf.* Nagel (1989): ‘The possibility of the development of conscious organisms must have been built into the world from the beginning. It cannot be an accident.’

backwards to understand how the world-historical potential for our development was actualized. This investigation will be privative and retrospective, working backwards from ourselves and our self-understanding by subtracting key properties to arrive at an understanding of preceding life forms. More complex animals, for example, can be understood as having a pre-linguistic understanding of things in the world of the kind which is frequently implicit in ourselves but which we can bring to conscious articulation when required. Such sentience can be traced back to simpler organisms which bring themselves into a state of satisfying equilibrium (through attraction and repulsion) which felt needs disrupt, and thence to individual cells which monitor their environmental boundaries.²⁷ This process may itself require the catalytic development of ‘autogens’ which construct those boundaries (*cf.* Deacon, 2012). While such a brief outline serves only to indicate a direction for the explanation of the historical emergence of the mental, it will have at least three important features. First: however far the process is traced back, at no point in this history is the world adequately characterized as merely physical: we are dealing with a history which must be understood as the stages of the development of our own being-in-the-world. The physical realm is always only one aspect of this development, so that at no point is there a need to cross the gap between the purely physical and the mental. Second: it will involve the irreducibly teleological concepts of self-organizing and self-maintaining organisms captured in the notion of autopoiesis,²⁸ concepts which cannot be adequately accommodated by the physical sciences (*cf.* Midgely, 2011). Third: it will recognize that while we can trace some of the necessary conditions of the emergence of consciousness, there is little chance of providing a non-circular set of sufficient conditions for its emergence.

Let us now situate this approach with respect to the other main parties in the discussion of the QEC. Clearly it is not a physicalist position, at least as far as epistemic reduction, eliminativism, or identity theory versions are concerned, although the possibility of the physical metaphysically necessitating the mental was left hanging as unattractively facing its own version of the QEC. Nor is this approach a standard substance or property dualism, each of which faces the QEC head on with unpalatable consequences. What of panpsychism? After all, does the claim that world history must be conceived of primarily as the

[27] *Cf.* the concept of autopoiesis explicated in Thompson (2007, Chapter 5).

[28] See Thompson (2007, pp. 140–9) for a discussion on the relation between autopoiesis and teleology.

process of development of agents in a life-world not imply that, prior to the emergence of life and consciousness, the fundamental pre-existing entities in the world have properties that are at least the precursors of consciousness? Here we must be careful. First, the notion that proto-consciousness is somehow smeared across microphysical entities leads to real difficulties as to how to combine these elements of consciousness into fully fledged experiences, let alone the subjects of such experience. But is this a price that has to be paid on the current account? The answer is no. It is important to be clear that despite the existential priority of physical entities, ontological priority must be accorded to the agent in the life-world. That is, what it *means* to be physical (now and 13 billion plus years ago) is to be one (quantifiable and objective) aspect of this agent/life-world totality, in abstraction from all subject-dependent and mental properties. At no point can we understand the agent/life-world totality — ‘Being-in-the-World’ — as an assemblage of physical components with some consciousness somehow stuck on, as those concepts are possible only as abstractions from that totality. The view we must take is that the world has always had the potential for development of Being-in-the-World, and one aspect of that is the causal structures of the physical realm.

Nor therefore is the account proposed here a form of monism, standard forms of which run into difficulties similar to panpsychism, except in so far as the focus on the agent/life-world as necessarily presupposed by our grasp of the physical and mental might be said to constitute a structured totalizing monism.

How finally does this account compare with the recent enactive, embodied, and phenomenological theories of mind with respect to the QEC? All of these theories focus on the notion of the engaged agent and over the last twenty years or so have tried either to resolve the QEC (e.g. O’Regan and Noë, 2001), to dismiss it as arising from the unreflective adoption of an objective stance which ignores the nature of experience which makes such a stance possible (e.g. Ratcliffe, 2007), or to ignore it and do something different instead: either neurophenomenology (e.g. Varela, 1997; Bitbol, 2012; Lutz and Thompson, 2003) and/or dynamic systems theory (e.g. Thompson, 2007; Deacon, 2012). It is arguable, though I cannot argue it here, that none of these accounts resolve the QEC.²⁹ Nevertheless, by developing an understanding of the relationship between the mental and the physical as abstractions from the agent engaging with the life-world, the

[29] See O’Connaill (2014b) for such an argument.

argument proposed here provides a philosophical underpinning for their further development.

7. Conclusion

The conclusion to be drawn from these considerations is that the QEC as normally understood assumes the ontological primacy of the physical world (from which consciousness must arise) without questioning the necessary conditions of the possibility of the central concept of the physical. Once this is done it shows that these conditions include the necessary presupposition of purposeful agents engaging with a life-world of affordances from which we can reflectively abstract both mental and physical properties. This offers an understanding of the physical and mental not as the alienated metaphysical categories whose relationship is the focus of the QEC, but as abstractions from the agent/life-world in the context of which alone their interrelations can be clarified.

Whilst physicalist approaches have been the prime target of this discussion, the main alternatives to physicalism currently in play, namely some form of dualism³⁰ or of panpsychism, are arguably in no better a position. Each of these approaches itself takes for granted an understanding of the physical and the mental without enquiring into the necessary conditions of the possibility of that understanding, and so fails to recognize the primacy of the notion of the agent or person, disruption in whose everyday engagements with the world allows for the reflective abstraction of those understandings and concepts. The problems encountered by each approach are then the consequences of trying to reconstitute the unity of the person by juxtaposing or combining the two abstractions.

The understanding of the emergence of consciousness developed here offers a sound transcendental basis for the continuing research in neurophenomenology, and a more productive approach to understanding its evolutionary development, than continuing to bang our heads on the various 'hard problems' of consciousness.

The hard problems that remain are those of fleshing out an approach of which this paper has provided only an outline, and of con-

[30] Including at least property dualism and traditional Cartesian substance dualism. Lowe (2008, Chapter 5) espouses a Non-Cartesian Substance Dualism which also prioritizes the subject as agent.

tinuing to argue that philosophical positions which ignore the necessary conditions of their own meaningfulness do so at their peril.³¹

References

- Apel, K.-O. (1980) *Towards a Transformation of Philosophy*, London: Routledge and Kegan Paul.
- Balog, K. (2012) In defense of the phenomenal concept strategy, *Philosophy and Phenomenological Research*, **84** (1), pp. 1–23.
- Bitbol, M. (2012) Neurophenomenology, an ongoing practice of/in consciousness, *Constructivist Foundations*, **7** (3), pp. 165–173.
- Brueckner, A. (2010) *Essays on Skepticism*, Oxford: Oxford University Press.
- Chalmers, D.J. (1997) Facing up to the problem of consciousness, in Shear, J. (ed.) *Explaining Consciousness — the Hard Problem*, Cambridge, MA: MIT Press.
- Chalmers, D.J. (2010) Consciousness and its place in nature, in Chalmers, D.J., *The Character of Consciousness*, New York: Oxford University Press.
- Chalmers, D.J. (2011) Verbal disputes, *Philosophical Review*, **120** (4), pp. 515–566.
- Crane, T. & Mellor, D.H. (1990) There is no question of physicalism, *Mind*, **99** (394), pp. 185–206.
- Deacon, T. (2012) *Incomplete Nature: How Mind Emerged From Matter*, New York: Norton.
- Denmett, D. (1991) *Consciousness Explained*, Boston, MA: Little, Brown.
- Dowell, J.L. (2006) The physical: Empirical not metaphysical, *Philosophical Studies*, **131**, pp. 25–60.
- Gibb, S.C., Lowe, E.J. & Ingthorsson, R.D. (2013) *Mental Causation and Ontology*, Oxford: Oxford University Press.
- Gibson, J.J. (1979) *The Ecological Approach to Visual Perception*, Boston, MA: Houghton Mifflin.
- Habermas, J. (1972) *Knowledge and Human Interests*, Boston, MA: Beacon Press.
- Heidegger, M. (1962) *Being and Time*, Oxford: Blackwell.
- Hempel, C. (1980) Comments on Goodman’s ways of worldmaking, *Synthese*, **45**, pp. 193–199.
- Husserl, E. (1970) *The Crisis of European Sciences and Transcendental Phenomenology*, Carr, D. (trans.), Evanston, IL: Northwestern University Press.
- Judisch, N. (2008) Why non-mental won’t work: on Hempel’s Dilemma and the characterization of the ‘physical’, *Philosophical Studies*, **140**, pp. 299–318.
- Levine, J.L. (2001) *Purple Haze: The Puzzle of Consciousness*, Oxford: Oxford University Press.
- Loar, B. (1997) Phenomenal states, in Block, N., et al. (eds.) *The Nature of Consciousness: Philosophical Debates*, Cambridge, MA: MIT Press.
- Lowe, E.J. (2008) *Personal Agency*, Oxford: Oxford University Press.
- Lutz, A. & Thompson, E. (2003) Neurophenomenology: Integrating subjective experience and brain dynamics in the neuroscience of consciousness, *Journal of Consciousness Studies*, **10** (9–10), pp. 31–52.
- McGinn, C. (1991) *The Problem of Consciousness*, Oxford: Blackwell.
- McGinn, C. (1997) Consciousness and space, in Shear, J. (ed.) *Explaining Consciousness — the Hard Problem*, Cambridge, MA: MIT Press.

[31] My thanks to Jonathan Lowe, Donnchadh O’Connail, and two anonymous reviewers for comments on earlier drafts of this paper.

- McLaughlin, B. & Bennett, K. (2011) Supervenience, in Zalta, E.N. (ed.) *The Stanford Encyclopaedia of Philosophy* (Winter 2011 edition), [Online], <http://plato.stanford.edu/archives/fall2009/entries/supervenience/> [30 March 2014].
- Midgely, M. (2011) Why the idea of purpose won't go away, *Philosophy*, **86** (4), pp. 545–561.
- Montero, B. (2001) Post-physicalism, *Journal of Consciousness Studies*, **8** (2), pp. 61–80.
- Montero, B. & Papineau, D. (2005) A defence of the *via negativa* argument for physicalism, *Analysis*, **65** (287), pp. 233–237.
- Nagel, T. (1989) *The View from Nowhere*, Oxford: Oxford University Press.
- Nagel, T. (2002) The psychophysical nexus, in Nagel, T., *Concealment and Exposure and Other Essays*, New York: Oxford University Press.
- Noë, A. (2009) *Out of Our Heads*, New York: Hill and Wang.
- O'Connell, D. (2014a) Grounding, physicalism and the explanatory gap, unpublished ms.
- O'Connell, D. (2014b) Phenomenology, objectivity and the explanatory gap, unpublished ms.
- O'Regan, J.K. & Noë, A. (2001) A sensorimotor account of vision and visual consciousness, *Behavioral and Brain Sciences*, **24**, pp. 939–1031.
- Papineau, D. (2001) The rise of physicalism, in Gillett, C. & Loewer, B.M. (eds.) *Physicalism and its Discontents*, Cambridge: Cambridge University Press.
- Poland, J. (1994) *Physicalism: The Philosophical Foundations*, New York: Oxford University Press.
- Ratcliffe, M. (2007) The problem with the problem of consciousness, *Synthese Philosophica*, **44** (2), pp. 483–494.
- Spinoza, B. (1985) Ethics Part III, in Curley, E. (ed. & trans.) *The Collected Works of Spinoza*, vol. I, Princeton, NJ: Princeton University Press.
- Stern, R. (ed.) (1999) *Transcendental Argument*, Oxford: Oxford University Press.
- Stoljar, D. (2005) Physicalism and phenomenal concepts, *Mind & Language*, **20** (5), pp. 469–494.
- Stoljar, D. (2009) Physicalism, in Zalta, E.N. (ed.) *The Stanford Encyclopaedia of Philosophy* (Fall 2009 edition), [Online], <http://plato.stanford.edu/archives/fall2009/entries/physicalism/> [30 march 2014].
- Taylor, C. (1978–79) The validity of transcendental arguments, *Proceedings of the Aristotelian Society*, **79**, pp. 151–165.
- Thompson, E. (2007) *Mind in Life*, Cambridge, MA: Harvard University Press.
- Trogon, K. (2013) An introduction to grounding, in Hoeltje, M., Schnieder, B. & Steinberg, A. (eds.) *Varieties of Dependence: Ontological Dependence, Grounding, Supervenience, Response-Dependence*, Munich: Philosophia Verlag.
- Varela, F.J. (1997) Neurophenomenology, in Shear, J. (ed.) *Explaining Consciousness — the Hard Problem*, Cambridge, MA: MIT Press.
- Wilson, J.M. (2006) On characterizing the physical, *Philosophical Studies*, **131**, pp. 61–99.
- Wittgenstein, L. (1953) *Philosophical Investigations*, Oxford: Blackwell.
- Worley, S. (2006) Physicalism and the *via negativa*, *Philosophical Studies*, **131**, pp. 101–126.