Why Physicalism?

William Seager

Abstract

Physicalism is the view that everything that exists is ultimately physical. It is the dominant metaphysics of nature in the current age despite facing a number of formidable challenges. Here I examine the reasons we have for believing in physicalism. It will turn out that the undeniable success of physicalism heretofore may in fact undercut the claim that physicalism deserves wholesale, even if provisional, acceptance. My argument stems from noting the disparity between 'ontological' physicalism—a doctrine solely about the nature of things—and 'epistemic' physicalism, a doctrine asserting the physical explicability of everything. The reasons we have for accepting physicalism necessarily stem from the history of success of epistemic physicalism. The problem of consciousness throws up a roadblock on this path toward physicalism, which then undercuts the grounds we have for endorsing ontological physicalism. This argument can be expressed in Bayesian form, which makes clearer the perhaps precarious position in which modern physicalism finds itself. I end with some more or less tentative suggestions for alternative metaphysical frameworks.

1 What is Physicalism?

The literature on physicalism: its nature, commitments, strengths and problems is staggeringly vast. I cannot hope to add anything definitive to this body of work in the way of settling any of these issues. What I want to do is to consider why physicalism is favoured amongst so many philosophers (and by no means only philosophers) and to ask whether its favoured status is really justified. I think the situation is a lot murkier than many think.

This project will require a characterization of physicalism. I hope to provide a fairly uncontroversial starting point and general overview but I can't avoid courting some controversy as the argument thickens.

Physicalism is a monistic metaphysics: it claims that there is only one basic kind of reality and it is physical in nature. The phrase 'kind of reality' is vague and ungainly; in the past materialist philosophers would have said there was only one kind of substance, or even more straightforwardly, only one substance: material, or physical, substance. But our grip on the idea that substance is the appropriate concept by which to describe basic reality has weakened. The rough idea remains clear enough: at bottom, everything is physical.

The nature of the physical is another very vexed philosophical issue (see for example Wilson (2006), Montero (2005), Strawson (2006), Stoljar (2001)) and may even be, as briefly addressed in the conclusion, a kind of mental construct rather than a simple

outgrowth of commonsense realism. I think physical reality is known in the first instance by ostension: we are in perceptual contact with aspects of the world which are paradigmatically physical. We begin there. But if known by ostension, the physical is *revealed* by science, most intimately by the foundational science of physics. Perceptually schooled intuition suggests something like a picture in which the physical is continuously extended, space filling and exclusively space occupying stuff. Unfortunately for intuition, science has revealed that the physical is much stranger than that and, so to speak, much less 'material'. This means that we must take a somewhat provisional attitude to the question and characterize the physical as whatever physics describes, or will end up describing, as underlying the ostensively familiar physical world. But the oddity of the physical as revealed by science also means there is serious difficulty understanding the relation between the physical as scientifically revealed and the familiar aspects of the observable world which beget the notion of the physical in the first place. This is a viciously hard problem in detail (see Belot and Earman (1997) for some of these details).

The problem is unavoidable though for, clearly, not everything is physical (in the scientific sense) under its usual description. On the face of things, innumerable features of the world are not obviously physical. Quite to the contrary, the physical is fundamentally non-chemical, non-biological, non-geological, non-meteorological, non-mental[1], non-social, non-political, etc. So, even if we grant that physics will eventually provide a comprehensive, complete and accurate account of physical reality, which is, for physicalists, all of reality, there will still need to be told a tale which relates the non-fundamentally physical to the fundamentally physical (i.e. what is revealed by physics).

Thus one can envision a grand view of the world which begins with the world as described by fundamental physics and ends with all the features we are familiar with in both ordinary experience and all the non-fundamental sciences. To a remarkable extent, we have this view already in our grasp (see chs. 1-3 of Seager (2012b) for a brief overview). There are untold numbers of known interconnections from the fundamental level to various non-fundamental features of our world which have been identified and explored. Looking at things from the reverse point of view, there are equally vast numbers of 'anchor points' where we can see, at least in general terms, how the non-fundamental springs from and depends upon the fundamental level. Any place where we don't yet see such interconnections or anchor points is a sore point, like a nagging splinter, but there remain very few such problematic areas.

In the abstract, physicalism thus demands that there be a dependence relation of the non-fundamental upon the fundamental. In order to sustain the claim of monism, this relation has to be pretty strong in at least two ways: logical and ontological. As to the first, the dependence relation must be of maximal logical strength: physicalism requires that it be absolutely impossible for two worlds to be identical with respect to the properties, laws and arrangement of the physical fundamentals and yet differ with respect to anything else. This relation is that of logical supervenience[2].

[1]This is the standard view, which equates mentality with other features which are not part of what physics takes to be fundamental. One of the main motivations for investigating panpsychist alternatives to physicalism arises from the apparent uniqueness of consciousness's failure to abide by the strictures of an acceptable form of emergence. This will be discussed below

[2]Logical supervenience is strong supervenience with the stipulation that the internal necessity operator be of maximal strength.

Logical supervenience is, in principle, consistent with non-monism if there is a maximally strong necessitation from the fundamental physical domain to some putative non-physical domain. For example, traditional epiphenomenalism can be made consistent with logical supervenience if the modal relation between the physical base and the supervenient mental state is 'bumped up' from the standard relation of causation to one of maximally strong necessitation. We might call this bizarre theory 'logical epiphenomenalism'. It is an interesting question exactly how logical epiphenomenalism is different from physicalism, if we allow that there are epiphenomenal features of physical reality, or that these are at least possible[3]. The natural answer is in terms of understanding how the emergent feature is constituted by the arrangement of physically fundamental features. I will say more about this below but on the face of it, a brute relation of maximally strong necessity between *distinct* domains seems extremely implausible. In fact, one might think that such distinctness is marked out precisely by possible modal variation.

The ontological constraint arises from respect for the picture of the world provided by fundamental physics. It appears that the world is, in some significant sense, made out of very small things. At least, the physical objects of familiar experience have parts which have parts... which have parts that (eventually) connect with the kinds of things described by fundamental physics. Presumably, then, the dependence relation we seek is some kind of complex relation of constitution.

It is the job of science broadly construed to work out the details of this constitution relation (or, more likely, the many constitution relations which will be involved in the long, ladder-like transition from the fundamental to the familiar), bearing in mind that the complexity of constituted entities will not permit anything like a full and completely transparent account. For example, we have a pretty good idea of how chemical kinds are constituted based on the principles of quantum mechanics, even though exact calculations from first principles will remain outside of our reach for all save the very simplest cases[4]. Philosophically however, we need only be concerned with the basic form of the constitution relation, which I suggest is something like the following (restricted for simplicity of presentation to a single propery, F), where C stands for a relation of constitution by which a is constituted by the z_i which stand in a relation Γ which 'generates' the property F:

[E] $Fa \rightarrow (\exists\Gamma)(\exists z_1, \ldots, z_n)(Cz_1, \ldots, z_n a \wedge \Gamma z_1, \ldots, z_n \wedge \Box_l(\forall x_1, \ldots, x_n, y)(Cz_1, \ldots, z_n y \wedge \Gamma z_1, \ldots, z_n \rightarrow Fy))$

It is important to bear in mind that a liberal interpretation of this formula is appropriate. There is no commitment to locality in the constitution relation (quantum mechanics

[3]When apparently epiphenomenal features arise in physical theory they are typically discarded as mere mathematical artifacts, but that is to take up a disputable philosophical stance. For example, only differences in electric potential are efficacious but absolute electric potential might be a real physical feature of the world. It is an interesting fact that frequently these mathematical 'artifacts' end up having genuine physical effects, as in the Aharonov-Bohm effect.

[4]Nonetheless, using various approximation techniques, it is possible to calculate macroscopic properties of substances from quantum first principles with more or less accurate results. For example: '[f]or small molecules in the gas phase and in solution, *ab initio* quantum chemical calculations can provide results approaching benchmark accuracy, and they are used routinely to complement experimental studies. A wide variety of properties, including structures, thermochemistry (including activation barriers), spectroscopic quantities of various types, and responses to external perturbations, can be computed effectively' (Friesner (2005)).

suggests that the properties of things are not determined fully locally) and similarly there is no implication that the correct constitution relation will support anything like part-whole reductionism (though it is compatible with it).

Obviously, [E] represents a form of emergence inasmuch as an object comes to possess a property which its constituents lack. Equally obvious is that physicalism will have to embrace some form of emergence. As noted, the properties deployed in fundamental physics are but a tiny fraction of the properties the world exemplifies. But there is nothing mysterious, mystical or transcendent about the emergence vouchsafed by [E]. No physicalist will be worried about this kind of emergence and in fact they should welcome its inclusion in the physicalist world view.

The logical strength of the inner necessity operator, $\Box_l$, is crucial here. It must possess the maximal modal strength of logical necessity (hence the subscript 'l') on pain of the intrusion of phenomena that are not suitably dependent on the physical fundamentals[5]. For example, if we were to replace logical necessity with mere nomological necessity the loss of logical supervenience would permit the existence of emergents which were not appropriately dependent on the fundamental physical state of the world. A change in the laws of nature could thus alter the distribution (or existence) of the emergent features. This would violate physicalism.

This latter view, that emergence should be understood in terms of a supervenience relation defined via nomological necessity is perfectly respectable and not unfamiliar. In essence, it was the view held by the so called British emergentists (see McLaughlin (1992)), notably Samuel Alexander (Alexander (1920)), Conwy Lloyd Morgan (Morgan (1923)) and C. D. Broad (Broad (1925)). I will label such a view radical emergence to contrast it with the conservative emergence enshrined in [E]. Independent of any merits or demerits of radical emergentism, it clearly does not meet the requirements of a robust physicalism[6].

Thus far we have been discussing ontological dependence. We have gleaned that ontological dependence is a synchronic relation which is non-causal. It is such that when X ontologically depends on Y then it is absolutely impossible for X to fail to exist if Y exists. In short, we can say that X ontologically depends on Y just in case Y provides the metaphysical ground for X. Physicalism can then be recast simply as the claim that everything ontologically depends on the physical.

The most frequent metaphor philosophers use to express what they mean by ontological dependence is a theological one: once God created the physical world, set the physical laws and the arrangement of the fundamental physical entities in the world, there was nothing left to do about the non-fundamental things. They would follow of necessity as a metaphysical 'free lunch' (another common expression used to denote

[5]Some would advocate a grade of necessity between nomological and logical that might afford the physicalist some more wiggle room. One example that is sometimes used to to illustrate this distinction is the metaphysical but non-logical necessity of the identity of water and H_2O. It is true that one cannot deduce any sentence with the term 'water' in it from sentences that only mention hydrogen, oxygen and quantum mechanics. But that is not what I mean by logical necessity. I simply mean that there are absolutely no possible worlds where there is H_2O but no water. I will discuss below a possible way to define metaphysical necessity which distinguishes it from a kind of logical necessity that is based on model theoretic considerations. For a classic discussion of the status of metaphysical necessity see Chalmers (1996), ch. 4.

[6]But it's worth noting that radical emergence is compatible with a weaker kind of physicalism. Radical emergentism can allow that the physical is the ontological base of the world out of which everything else emerges.

ontological dependence). If some physically non-fundamental entity was a radical emergent, then God's job would not be done with the creation of the physical. God would have to add in the 'laws of emergence' in order to ensure the generation of such emergent features.

There is, however, another quite distinct dependence relation that we should also consider, that of epistemological dependence. What I mean here is the dependence of understanding some aspect of reality upon understanding some other aspect of reality. There is a very famous saying of Christian Dobzhansky: 'Nothing in biology makes sense except in the light of evolution' (Dobzhansky (1973)). Though perhaps somewhat overstated, the remark expresses well the idea of epistemological dependence, albeit one of an extreme form.

We can define absolute epistemological dependence thus:

> X is *absolutely* epistemologically dependent on Y iff it is impossible to understand X except via an understanding of Y

There are a number of plausible candidates for domains that are absolutely epistemologically dependent on other domains. It is, for example, surely impossible to understand politics on Earth without having an understanding of our distinctively human psychology. Absolute epistemological dependence generates a necessary condition for understanding but makes no claim about sufficiency. But it we take the union of all the domains on which a given subject matter is epistemologically dependent, it is likely we will have specified a sufficient base for understanding that subject matter.

Some domains are epistemologically independent, even if in the actual world these domains are constitutionally related. The abstract theory of computation (a part of mathematics) is epistemologically independent of the theory of transistors and electrical circuitry, even though all our computing devices are made out of carefully organized systems of 'circuit elements'.

Furthermore, some domains can be understood independently but can also be understood via an understanding of other domains. For our purposes, the important form of this weaker relation is one I will call reductive epistemological dependence, defined as:

> X is reductively epistemologically dependent on Y = it is possible to understand X via an understanding of Y

All cases of scientific reduction (broadly construed) will generate a relation of reductive epistemological dependence. For example, although the principles of thermodynamics were independently discovered and understood, it is also possible to understand thermodynamics from an appreciation of statistical mechanics. In fact, SM gives tremendous insight into the nature of thermodynamic principles and a deeper understanding of how and why they obtain so universally. Chemistry too was developed to a sophisticated level without any understanding of the physical principles underlying it, but it is possible (at least in principle) to understand chemistry on the basis of fundamental quantum mechanics.

A more or less outrageous philosophical thought experiment provides a kind of test for the existence of reductive epistemological dependence: try to imagine a capacious mind with access to everything about the reducing domain, and consider whether such a mind could on that basis figure out, or come to understand, the reduced domain. It is perfectly legitimate to further imagine providing this mind some help in defining

the concepts characteristic of the reducing domain to get around the purely logical fact that novel concepts cannot be 'deduced' out of old ones. Reductive epistemological dependence should not be thought of as pure logical deduction from nothing but the resources available in the reducing domain. Could such a hypothetical mind (which C. D. Broad called a 'mathematical archangel') figure out the chemical properties of water given sufficient knowledge of the quantum structure of hydrogen and oxygen? Given the extant results of ab initio chemistry accomplished by our intellectually puny human scientists (compared, that is, to the archangel), it seems that our capacious mind could indeed derive chemical understanding on this basis.

Note my caveat that the kind of reduction involved here should be 'broadly construed'. I do not think that reductive epistemological dependence requires anything like a full deduction of the target domain from knowledge of the reducing domain. Sheer complexity, if nothing else, will block any such full and complete deduction. Reductive epistemological dependence only requires that the target be made intelligible in terms of the reducing system.

The interplay between ontological and epistemological dependence will, I think, help us understand the nature and prospects of physicalism.

2 Evidence for Physicalism

If one steps back from any preexisting metaphysical commitments to physicalism, one might wonder what is the source of its widespread acceptance. I want to present four arguments (or at least motivations) that seem important:

1. Unparalleled scope, scale and explanatory power of physicalist metaphysics
2. Induction
3. Intrusions from below
4. Methodological Integration

The first argument needs little elaboration. It is an undeniable fact that vast swathes of the world have fallen under the explanatory project of physicalism. We now possess the outlines, and in may areas far more than outlines, of a truly grand metaphysical edifice which encompasses the entire spatial extent and history of the observable universe. This physicalist picture of the universe perhaps comes close to fulfilling Wilfred Sellars's general characterization of the project of metaphysics: '...to understand how things in the broadest possible sense of the term hang together in the broadest possible sense of the term' (Sellars (1963), p. 37).

A clear expression of the power of the physicalist world view, as well as some of its defenders fervency, can be found in these remarks of the physicist (though one quite philosophically inclined) Sean Carroll: 'The laws underlying the physics of everyday life are completely understood...All we need to account for everything we see in our everyday lives are a handful of particles—electrons, protons, and neutrons—interacting via a few forces—the nuclear forces, gravity, and electromagnetism—subject to the basic rules of QM and GR...' (Carroll (2010)).

There is no gainsaying that the physicalist picture is intellectually compelling in its scope, explanatory power, metaphysical simplicity and ontological elegance.

The second argument might be labeled 'the optimistic induction'[7]. It is based simply on that fact that the scientific enterprise has, for over four hundred years now, enjoyed an ongoing pattern of success based, in important part, on the underlying presupposition that there is a physicalist account of all phenomena. Of course, there have been many times when this success appeared threatened or stalled in the face of recalcitrant elements of nature. But these have quite uniformly eventually yielded to a physicalist understanding and integration with the larger scientific picture of the world.

Here is one instructive example. At the beginning of the 19th century it was far from obvious that life, and in particular organic chemical compounds, would be susceptible to standard materialistic chemical understanding. Vitalism was a respectable scientific doctrine (and would remain so for a long time, with a gradually decreasing following). We can regard vitalism as a sort of radical emergence which claims that some genuinely novel element of reality comes into being when certain chemical substances form an organic system.

In 1824 it appeared that radical emergence and vitalism might actually be empirically verified. Around 1800 chemistry was already very well established as a science and struggled with the problem of identifying the elemental constituents of various substances, especially the intrinsically complex organic compounds. In 1824, a pair of promising young chemists, Justus von Leibig and Friedrich Wöhler, managed to identify two different substances, cyanic acid and fulminic acid[8]. These acids have quite distinct chemical properties and could be unambiguously categorized and differentiated in chemical terms. Astonishingly, Leibig and Wöhler discovered that each acid had identical elemental constituents in identical ratios. Chemical orthodoxy at the time held, in effect, that chemical properties supervened on elemental composition and ratios thereof. Something of a crisis ensued and both Leibig and Wöhler regarded the other with suspicions of incompetence. But when they collaborated on a more careful analysis the same result was achieved: fulminic and cyanic acid were, apparently, chemically 'identical'.

Of course, this was not a victory for vitalism or radical emergence. The solution was to take the physicalistic picture yet more seriously. By embracing the idea of literal spatial atomic structure, it was possible to explain the difference between fulminic and cyanic acid in terms of the arrangement of the constituents, which can differ even if the elemental composition of both was the same. It turned out that fulminic and cyanic acid were what are called isomers: identical composition but distinct spatial structure.

There are untold numbers of similar roadblocks successfully circumvented, always within the general physicalist framework. The optimistic induction is the inference that in the face of such long term and extensive success, the scientific metaphysics will be completed. It would take a brave person to bet against this trend.

However, for our purposes we need not dwell on this long history of success. What is important to note is that throughout this four hundred year hot streak there has been one constant trait. The physicalist viewpoint has advanced via the revelation, often hard and slowly won, of the mechanisms of ontological dependence. That is to say, via the exhibition of reductive epistemological dependence. All of these successes have shown us how we could understand some (relatively) 'macro' phenomenon via an understanding

[7]In contrast to the pessimistic induction much bruited by philosophers of science, which is the evident historical fact that all scientific theories save our current ones have turned out to be false. The optimistic side is that this woeful history of 'failure' has never revealed a problem with the physicalist presupposition behind the successive replacement of one theory by another.

[8]For more on this story see Brock (1993), ch. 5.

of some (relatively) 'micro' structures and processes. There is a very tight connection between ontological and epistemological dependence: the former's plausibility depends on the exhibition of the latter.

The third argument for physicalism—intrusions from below—is rather more subtle, but telling. We find in nature what is evidently a hierarchy of structure which is of a very special kind. Many levels in this hierarchy sustain descriptions in terms of their own, as it were proprietary, laws. These levels have a kind of autonomy which permits us to ignore levels below. This is a very robust form of emergence. But it does not fall outside the realm of conservative emergence in any way that threatens the physicalist viewpoint. How do we know? Even in the absence of a complete understanding of the mechanisms of cross-level ontological dependence, there is a telltale sign: intrusions from below. We find that the autonomy of a given level in the natural hierarchy is broken from time to time, the laws of that level fail to hold, because of some effect of a lower level.

There are many examples of such intrusions from below. In chemistry, it is a rule of thumb that isotopes share chemical properties but this is not a perfect rule. If you drink heavy water for a while, you will sicken and die. The extra mass of the deuterons in heavy water subtly changes reaction rates and prevents cellular metabolism from proceeding normally.

A more rarefied example is the mighty domain of thermodynamics, whose power to order the world is everywhere visible. But while the laws of thermodynamics are exceptionless as written, they are subject to statistical fluctuation and the very slight possibility of reversal. This is because the 'implementation' of thermodynamical properties is via a vast system of micro-states which can, in principle, be ordered so as to lead to violations of the letter of thermodynamic law. In a famous thought experiment, Josef Loschmidt showed that Boltzmann's 'deduction' of the second law of thermodynamics from statistical mechanics was flawed, since it was evidently possible that there could exist a dynamically reversed version of any micro-state (multiply all momentum values by -1). Such a system would exhibit thermodynamically impossible behaviour, as in the water in a bathtub suddenly sorting itself out so that the hot water was at one end and the cold at the other. It is thus possible for the micro-state to intrude into the thermodynamical level[9].

A third example is from biology. Population dynamic equations are mathematically continuous but, obviously, any real biological population is composed of discrete individuals. This leads to anomalies called lattice effects, which are intrusions from below interfering, so to speak, with the representation of the population as a continuous variable (see Henson *et al.* (2001)).

Any number of similar examples could be found across the sciences. The positive upshot for physicalism is that the obvious explanation for the prevalence of intrusions from below is that the hierarchy of levels in nature exemplifies a system of ontological dependence, exactly as predicted by physicalism.

The final argument for physicalism is not really an argument at all, but rather a kind of stance taken by a large number of philosophers. This is a stance of methodological solidarity with the sciences wherein philosophy's role is that of handmaiden or underlabourer to science (see Ney (2008) for an interesting discussion). Overall, given that the foundational science of physics is taken to provide a pretty good account of the most

[9]For more on the Loschmidt-Boltzmann controversy see Sklar (1993)).

basic structure of the world, physicalism presents a general picture of the world which is conceptually simple. The physicalist viewpoint is ontologically pure, so to speak, unsullied by worrisome extraneous (or even supernatural) elements that threaten the accuracy of the scientific picture of the world. And, not to be underestimated, the physicalist metaphysics allows philosophers to side with science, to engage in a project that is continuous and cooperative with as well as deferential to science.

While no argument for the truth of physicalism, I think a desire for affinity with science is a powerful motivator for philosophers. It is quite natural nowadays to fall into the thought that in the search for the nature of ultimate reality, physics is the forefront discipline[10]. If so, the project of constructing a general metaphysics on the rich base science provides is compellingly attractive and, enticingly, apparently almost complete. Furthermore, even if the foundational role of science, especially physics, is accepted a host of rich philosophical problems remain in sorting out the physicalist metaphysics and its connection to the familiar world of everyday experience. It is certainly a noble, demanding and undeniably worthwhile philosophical project to see how far the physicalist metaphysics can be taken.

But, we should be wary and vigilant that the allure of science integration and the joy of metaphysical construction does not get taken too far.

3 The Barrier of Consciousness

Notwithstanding the foregoing, for a long time now the forward march of physicalism has been impeded, if not halted, by the phenomenon of consciousness. Of course, there is a vast amount of evidence linking consciousness to the brain. It is now possible to 'read off' some states of consciousness by observation of brain activity (see e.g. Owen (2008); for a general discussion of this evidence see Seager (2012b), ch. 4) and even in some small measure to reconstruct the nature of visual experience via real time MRI measurements (Nishimoto *et al.* (2011)). This is remarkable and even somewhat disturbing work, unmistakably signaling future abilities in the realm of 'mind reading' the social and ethical implications of which should be carefully reflected upon. And such work is certainly part of what is needed to bring consciousness within the fold of physicalism.

But if the sketch given above of how physicalism progresses is correct, this sort of evidence is far from sufficient to complete the physicalist's task. What is needed is the exhibition of the epistemological dependence of consciousness on the properties and arrangements of the physical processes which underlie or implement it.

The conspicuous lack of even a hint of how consciousness could be reductively epistemologically dependent on the physical is the main target of all the classic anti-physicalist arguments concerning consciousness. These arguments are so well known that there is no need to discuss them in detail here, but to set them in the current context let us briefly recall the Big Three: Thomas Nagel's 'what it is like argument' (Nagel (1974)), Frank Jackson's 'Mary argument' (Jackson (1982)) and the Descartes /

[10]Such an idea is exploited by science lobbyists, as illustrated by a remark from the United States LHC Communications Task Force, a body which serves to advocate American involvement at the European LHC facility. The first strategy towards this goal listed in their report is to 'promote recognition by key audiences of the value to the nation of particle physics, because of... its unique role in discovery of the fundamental nature of the universe' (Banegas *et al.* (2007)).

Saul Kripke / David Chalmers 'modal-conceivability argument' (Descartes (1641/1985), Kripke (1980), Chalmers (1996))[11].

Nagel quite explicitly highlights our complete lack of understanding of how the physical operation of the brain could generate consciousness: '[w]e do not have the beginnings of a conception of how it [physicalism about consciousness] might be true' (Nagel (1974), p. 177). Nagel is officially agnostic about the truth of physicalism, or even leans towards accepting it[12], but takes it for granted that absent a plausible route towards establishing reductive epistemological dependence, arguments in favour of a physicalist solution to the mind-body problem are just 'sidestepping it' (p. 180).

One can regard Jackson's argument about the physically omniscient neuroscientist from the same point of view. It is because of the lack of any glimmering of how consciousness could epistemologically depend on brain activity that the conclusion that Mary would not know what it is like to see red prior to her first experience of it is so intuitively appealing. If we transform the argument to one where Mary knows all about physics but nothing about chemistry its intuitive pull completely evaporates. The claim that knowing all about the physics of hydrogen and oxygen would leave Mary in the dark about, say, the boiling point of water is ludicrous. It is of course true that most chemical properties are epistemologically remote from their physical basis in terms of complexity and computational intractability, but it is enough for us to see in some intelligible and reasonably clear way how chemical properties could be epistemologically dependent on the physical. This we can do for chemistry but we have no clue about any such linkage from brain to consciousness. Sanguine physicalists will plead for more time: once sufficient knowledge of the brain has been accumulated, epistemic transparency will ensue. I am more inclined to the view that the pattern of our ignorance here shows that the problem of consciousness represents an entirely new kind of problem, never yet faced in the advance of physicalism.

Finally, what funds the conceivability argument whose conclusion is that there is a modal gap between the physical and consciousness is, again, the patent absence of any sense of how consciousness could be reductively epistemologically dependent upon the physical. While obviously there is no guarantee that such an absence ensures a real modal gap, it clearly opens up some space for this conclusion. To compare this situation to the chemical once again, it is flat out impossible for the physical to be arranged as it is around here (i.e. current state and laws of fundamental physics in place) and for water to not boil at 100° C. We find ourselves unable to make the leap to such an 'impossible world' just because we have a reasonable idea of how chemical properties are reductively epistemologically dependent on the physical.

So all these arguments can be viewed as traveling from a sense of failure in the project of exhibiting epistemological dependence to doubts about physicalism. But so what? Why should these doubts even arise? Physicalism is about ontological dependence. Not, in the first instance, epistemological dependence. And the classic arguments are in themselves not uncontroversial and place a heavy load on certain philosophical intuitions. Furthermore, it might well be that the failure of epistemological dependence is in some way itself to be expected or is at least explicable without threatening ontological

[11]I am running roughshod over important differences in the Descartes-Kripke-Chalmers arc of argumentation, but these will not matter to my discussion here.

[12]At least, that was true at the time Nagel wrote the famous bat paper and in Nagel (1986); he seems to have definitively rejected physicalism in his latest work, Nagel (2012).

dependence.

I think to the contrary that the three arguments really do point to a major problem for physicalism. Although it is true that the linchpin claim of physicalism is that of ontological dependence, if we ask for the source of evidence in favour of ontological dependence, there is only one answer: exhibition of reductive epistemological dependence. The classic arguments strongly emphasize that there is no glimmer of any viable relation of reductive epistemological dependence of consciousness on the physical, something which by now even a large number of committed physicalists accept. The whole dialectic is in a peculiar position.

4 The Burden of Proof

Does the long history of physicalist success make physicalism the 'default' position? I think this is a delicate issue. The long history of success in the construction of the physicalist world view has uniformly proceeded by integration or assimilation of phenomena into that view. This integration has been achieved via development of more encompassing relations of broadly reductive epistemological dependence, allowing us to *understand* more and more of the world in terms of the physicalist metaphysical picture. This history has never been and could not be the simple exhibition of ontological dependence.

If we imagine someone arguing by induction from the history of success we have to be careful about what the induction would actually be over, ontological or epistemological dependence. What the history of the success of physicalism suggests is a principle like the following:

> If physicalism is true then any phenomenon will stand in (at least) a reductive epistemological dependence relation to the physical

This is the proposition for which we have historical inductive evidence. The really important claim of ontological dependence follows on from successful exhibition of relations of epistemological dependence.

On the other side, how would a failure of ontological dependence be shown? One way, of course, is to show that some target phenomenon actually varies independently of the physical. Such independence, whether theoretically based or empirically evidenced, is what ultimately justifies our catalog of distinct fundamental *physical* entities and properties. In the case of consciousness, we do not expect to find independent variation, at least not of any clear cut kind, for at least two reasons. One, we already have abundant evidence for robust brain-consciousness links. If we accept a law-like relation between consciousness and corresponding brain states, we would not expect to ever find intra-world variation. Two, by its nature, hypothetical variation in the physical-consciousness linkage is invisible. Suppose that half of human beings saw colours differently but in such a way that all structural or relational intra-colour properties were preserved (whether the famous thought experiment of colour inversion could meet this condition is difficult question—see Hardin (1988) for a classic discussion of the difficulties here—but that does not matter as long as some colour variation could do so). Such variation in colour experience would be both behaviourally and neuroscientifically undetectable[13].

[13]One might wonder, if we are looking for variation in consciousness, why we could not find it intra-subjectively. I would surely notice if my colour vision suddenly inverted. Again, there is no reason to doubt

For, of course, there are famous relations of metaphysical independence that can mimic the appearance of dependence, viz. the relation of causation and common cause parallelism. Such impostors can be eliminated if we can exhibit reductive epistemological dependence. So exhibition of epistemological dependence is a *prima facie* requirement for extension of physicalism.

Persistent failure to discover epistemological dependence obviously points to the lack of such dependence. This in turn surely points towards a failure of ontological dependence. The strength of physicalism is its long history of success, but this is a history of revelation of patterns of epistemological dependence. Thus—somewhat paradoxically—in the face of the stubborn reluctance of consciousness to be integrated into the scientific metaphysics via the normal route of epistemological dependence, the long history of success of physicalism is evidence *against* it.

We can express this somewhat peculiar situation semi-formally in a probabilistic framework. What we have established is that the long history of physicalist success inductively implies that

$Pr(E|P)$ is very high,

where E=consciousness is reductively epistemologically dependent on the physical; P=physicalism is true. We can express this conditional probability in a mathematically equivalent way, as:

$Pr(E|P) = Pr(E\&P)/Pr(P)$.

This latter expression can then be rewritten as:

$Pr(E\&P)/Pr(P) = \frac{Pr(E)\times Pr(P|E)}{Pr(P)}$.

The somewhat complicated ratio above is full of interesting probabilities, especially the denominator which give us the chances that physicalism is true. But, what is the value of $Pr(P|E)$?

To get a handle on evaluating this, let us suppose that we had in hand a viable relation of reductive epistemological dependence of consciousness on the physical. Given the general success of the physicalist project it is plausible that consciousness is the *only* phenomenon for which we lack (at least an intelligible outline of) the relation of epistemological dependence. If so, then on this assumption physicalism is almost sure to be true, that is, $Pr(P|E)$ is extremely high.

Let us now consider the value of $Pr(E)$. The value of this probability appears to be close to zero. Intuitively, there seems to be no way for there to be anything like the standard sort of reductive epistemological dependence in the case of consciousness. The problem of phenomenality or 'what it is likeness' seems to be utterly different than previous problems faced in the expansion of the physicalist world view (of course, that may be mere appearance—perhaps tomorrow the scales will fall from our eyes and the epistemological dependence of consciousness on the physical will become transparent). This intuition is developed, deepened and bolstered by the three classic arguments.

It is also striking that a large number of modern physicalists agree that there will never be anything like the standard physicalist integration of consciousness into the physicalist picture of the world (see e.g. Loar (1990), Levine (1983), Papineau (2006), McGinn (1989)).

---

that there are nomological relations between the structure and processes within my brain and my states of consciousness.

Now, given that $Pr(E|P)$ is very high, then in our final expression we can see that $Pr(P) \approx Pr(E)$, which is to say, $Pr(P)$ is very low. We can diagnose the situation as follows: the very great success which physicalism has had in assimilating all phenomena (save consciousness) within reductive epistemological dependence means that any really serious roadblock casts really serious doubts on the truth of physicalism. And consciousness appears to be exactly this kind of roadblock.

5 Massaging the Values

The mathematics in the foregoing is trivially correct and so the only possible response on behalf of physicalism is to modify some of the 'input' probabilities we used. For example the position of mysterianism, whose chief exponent is Colin McGinn (see e.g. (1989))[14], which claims that it is our own intrinsic intellectual weakness that prevents us from understanding how consciousness is epistemologically dependent on the physical, can be viewed as arguing that $Pr(E)$ is, in some important sense, actually not low. Mysterians are so committed to physicalism they think there simply *must* be a suitable relation of reductive epistemological dependence of consciousness on the physical. It is just that our puny minds are incapable of coming up with that relation or, perhaps, even of understanding it if it were provided to us.

It is, of course, quite plausible that there are intrinsic limits to human intellectual abilities which might well put some domains beyond our ken. But it is extremely strange that there is only one such domain: consciousness. Why in the world would there be only and exactly one phenomenon that resists entrapment in the folds of epistemological dependence? Other areas of philosophical interest can seem to resist physicalist assimilation, but for them to present genuinely additional problems, they must retain their recalcitrance on the assumption of physical *and mental* stability. For example, there is a problem of naturalizing ethics. But there are many reasonably plausible accounts of how ethical values depend on the distribution of physical and mental features throughout the world. It seems hard to conceive of modal variation in ethical facts given identical physical arrangements and sentient responses. Only in the domain of consciousness does there remain a sense of possible variation in the face of physical (qualitative) identity[15]. It seems more likely that it is the nature of consciousness itself that creates our difficulty in understanding how it could be physical rather than the reverse situation wherein it is an unsolvable problem in understanding the physical nature of consciousness that makes the latter seem so odd.

A more reasonable, and much more widely adopted, approach which has the effect of lowering the value of $Pr(E|P)$ is to embrace a 'dual pathway' model of epistemological access to states of consciousness. This is usually explained in terms of our possessing

[14]The label of 'mysterianism' was first bestowed on this doctrine by Owen Flanagan (1991).

[15]McGinn has attempted to extend the claims of mysterianism beyond the realm of consciousness to encompass a number of classically intractable philosophical problems (see McGinn (1993)). Leaving aside the plausibility of these attempts, a goodly number seem clearly to depend upon the problem of consciousness. These include the issue of free will, the self and meaning. Consider the problem of free will. Absent the distinctive experience of apparent freedom in our choices, various naturalizing theories of the will seem quite acceptable. It is open to the physicalist to deny that there is any such thing as freedom of the will as philosophers take it. This may be false, but it does not seem in any way incoherent, as is the denial of consciousness.

two distinct set of concepts: standard physical concepts of the brain and its states, available from the third person standpoint and in addition a set of 'phenomenal concepts' which are applied first-personally to experience (for various attempts along these lines see Loar (1990), Rey (1993), Papineau (2006)). The reference of both sets of concepts is the physical basis of consciousness. The appearance of an epistemic gap or conceivable modal variation arises from the cognitive difference between these kinds of concepts. The crucial difference is that the phenomenal concepts are applied to experience 'directly' whereas the standard concepts are applied via some indirect epistemic route (for example, via examination of a brain scan or from some other sort of instrumentation, as in the famous, albeit for the present fictional, auto-cerebroscope perhaps first deployed by Paul Meehl (1966)).

In my view, this approach faces a severe difficulty which arises when we consider how phenomenal concepts are applied to experience. It seems that their application depends upon an appreciation of some feature of experience which can be regarded as something like a 'mode of presentation' of whatever property is the reference of the concept, which might well be a physical property. As in classic Fregean cases of distinct thoughts about the same thing, there must be a presentational difference to the subject which sustains the idea that there are two things under consideration. It is natural to understand such presentational differences in terms of properties of the (single) object which are, of course, distinct properties (the appearance and demeanor of Clark Kent versus that of Superman for example). Thus, in the case of consciousness, experience may present itself as either phenomenally conscious or as physical (via instrumentation). The properties of phenomenal consciousness in virtue of which we apply our phenomenal concepts are distinct properties from neurological properties, or at least they appear to be distinct, and so our problem re-appears at the level of modes of presentation. This objection to the phenomenal concepts strategy has been deeply explored by Stephen White (see White (1986, 2010)). I think it provides powerful addition grounds for respecting the intuitive rationale for the classic anti-physicalist arguments.

One possible reply for a defender of the phenomenal concepts strategy for lowering $Pr(E|P)$would be to take the application of phenomenal concepts to be the product of a brute recognitional capacity which thus does not depend on appreciation of any feature of experience. In general, the existence of such capacities does not seem particularly hard to accept. A famous philosophical example is the putative ability of chicken sexers to determine the sex of chicks by feel without having any awareness of what is about the chick that grounds the discrimination; discussed—to a different end—in Armstrong (1968), pp. 114ff.. The by now familiar phenomenon of blindsight is a more striking and highly relevant example. As is well known, certain sorts of brain damage to the visual centres of the brain can create scotomas in the the visual field within which subjects claim they have no visual experience. Despite this, if forced to guess whether, say, a light is off or on in this region of the visual field, subjects will answer correctly. Though highly unusual, and of course pathological, blindsight abilities of this kind can be regarded as pure recognitional capacities.

The problem with this approach is then pretty clear. It is wildly implausible that when we apply phenomenal concepts we do so in the absence of any 'source material' in experience on the basis of which we categorize phenomenal consciousness. Or, to put it another way, if application of phenomenal concepts was via such pure recognitional capacities, then this would be evident to us. I'm not sure this example is perfect, but compare how you know how your limbs are currently arranged (without looking!)

with how you know what colours your are experiencing[16]. I know both, but the former knowledge does not seem to be mediated (in general) by any particular quality of my experience (save when my limbs are in unusual and uncomfortable positions or have been motionless for a long enough time to generate pain), but my awareness of colours is obviously vividly phenomenological. The psychological literature is replete with examples of neurological disorders that feature what might be called knowledge without awareness, as in blindsight but there are many others[17]. It is of course striking that what is missing in these cases is specific sorts of consciousness despite the presence of certain recognitional capacities.

If one takes the recognitional capacities approach to its logical conclusion, consciousness becomes a kind of illusion. On this view, there is no phenomenal experience, but we possess a rich and complex set of concepts which have no genuinely referential application but instead describe a non-existent world in a proprietary manner[18]. Recognitional capacities trigger the application of these concepts which over the long span of human cognitive, intellectual and cultural development discursive thought has elaborated into a structure which supports a rich but entirely delusive system of beliefs. In terms of what we think consciousness is from *within* this system, we are actually no more conscious than rocks.

A very clear expression of this view is provided by Daniel Dennett, who in *Consciousness Explained* describes:

> ... a neutral method for investigating and describing phenomenology. It involves extracting and purifying texts from (apparently) speaking subjects, and using those texts to generate a theorist's fiction, the subject's heterophenomenological world. This fictional world is populated with all the images, events, sounds, smells, hunches, presentiments, and feelings that the subject (apparently) sincerely believes to exist in his or her (or its) stream of consciousness (Dennett (1991), p. 98).

I hesitate to ascribe this view to Dennett since his writing is often ambiguous between a position that is solely devoted to debunking certain perhaps dubious philosophical notions, such as that of qualia, which purport to characterize conscious experience and a position which entails the wholesale denial that there is anything even remotely like phenomenal consciousness in the world. The former attacks a straw man. The latter position is surely absurd. The problem of consciousness does not revolve around descriptions of consciousness but around the simple fact that conscious beings are *presented* with the world, and themselves, in a special way quite different from the causal and information laden reactions of more ordinary physical objects.

The idea that presence is a fictional object seems too wildly implausible to be taken seriously, yet seems to the be the natural upshot of the pure recognitional capacities

[16]For a detailed investigation of the relevance of various blindsight thought experiments to the problem of consciousness, see Siewert (1998).

[17]A particularly fascinating example is discussed in Goodale and Milner (2004). The unfortunate subject, who suffered carbon monoxide induced brain damage, is able to perform a number of complex perceptual tasks without awareness.

[18]A vaguely analogous situation might be the rich set of concepts developed by Christian thinkers in the late middle ages to describe the occult world of demons, angels and witches. The familiar scientific example of phlogiston can serve as another illustration. A reasonably complex and empirically successful theory was developed around the notion of this non-existent substance.

interpretation of phenomenal concepts.

If dual access accounts fail to lower the value of $Pr(E|P)$ perhaps a more radical effort is needed. It seems possible, in principle at least, for a physicalist to hold that it is simply a brute, primitive fact that consciousness supervenes with maximal logical strength upon the physical. Since this involves the outright denial that there is any intelligible connection between the physical and consciousness, there is no possibility of their being any reductive epistemological dependence of consciousness on the physical. Thus the value of $Pr(E)$ is exactly zero and hence $Pr(E|P)$ is also zero.

Despite this technical success, the brute necessity option should be deeply unsatisfying to the physicalist. It entails abjuring any prospect of completing the physicalist view of the world and replaces it with little more than the bare assertion that consciousness is ontologically dependent upon the physical.

This approach also leaves the physicalist in the unhappy position of having to admit that consciousness remains a unique feature of the world: the only one for which ontological dependence is inexplicable in principle.

Furthermore, it forces the physicalist into what I think is a very uncomfortable position about the nature of necessity. There is no reason to posit such brute necessities and they are on the face of it bizarre (see Chalmers (1996), pp. 136-43). The absence of any connecting link between two domains shows that modal variation is possible between them. Certainly, in the case of physicalism positing a brutely necessary dependence of consciousness on the physical there is an obvious air of the ad hoc. In no other area do we need to postulate such brute necessities.

We can see how strange the posit of brute necessities is if we compare it to more familiar modal brute facts. The most basic laws of physics and the most basic physical quantities are those which do not depend on other laws and quantities. Of course, we always search for more fundamental laws and principles so the 'catalog of brutality' (so to speak) is provisional, but the fundamental features of the world, whatever they may by, are brute facts, inexplicable and to be accepted with 'natural piety'[19]. For example, the standard model of physics, our most comprehensive and extremely successful fundamental physical theory—see the quote from Sean Carroll above—contains eighteen parameters whose values are, theoretically speaking, arbitrary—they must be measured (see Cahn (1996)). One could postulate that these values are more than nomologically necessary, that they are in fact absolutely necessary but clearly in the absence of any grounds to support this, such a postulation is entirely unjustified. Instead, these parameters mark out a space for modal variation. Such variation is the subject of much lively discussion in fact (see e.g. Barrow and Tipler (1988)) which has led to fascinating cosmological insights.

It is important to forestall a possible confusion here. It is sometimes maintained that the problem of integrating consciousness into the physicalist picture is transformed when we appreciate that physicalists seek to *identify* consciousness with certain physical states (see Block and Stalnaker (1999). In our terms, if we identify A with B then there is no question of establishing a relation of reductive epistemological dependence of B upon A (or vice versa). For example, if the morning star is identical with the evening star then it makes no sense to ask for an illuminating account of why this identity

[19]This is the phrase of Samuel Alexander, a radical emergentist, who wrote that natural piety is the attitude of the scientific investigator 'by which he accepts with loyalty the mysteries which he cannot explain in nature' , (Alexander (1922), p. 609).

holds. In such a case, the task which makes sense is to explain why the identity was not recognized, or why it appears the identity does not hold. The phenomenal concepts approach considered above is, of course, ideal for this task.

This line of argument fails here, however, because it neglects the fact that whatever physical state is to identified with consciousness, it will be a very complex, multi-component, state. It will be a state constituted out of simpler, ultimately fundamental, physical entities. Thus the question will persist how an assemblage of just these fundamental physical components suffice to generate consciousness. To put it another way, we still need a story which makes intelligible the conservative emergence of consciousness from the selected physical basis. The bare physical story of how the fundamental parts fit together is not automatically going to be the story of how consciousness is generated.

It would be different if the physicalist wanted to identify consciousness with some fundamental physical feature. But, besides the fact that no physicalist would have such a perverse desire, there is no fundamental feature in the physical picture which suggests itself. On the other hand, such a suggestion is congenial to the panpsychist who argues that consciousness, in some presumably unutterably primitive form, must characterize physical reality at the most basic level. This line of thought is what prompts Galen Strawson (2006)to call his panpsychist view 'real physicalism'. But it is a core element of the physicalist ethos under consideration here that at the most basic level the physical is thoroughly non-mental.

6 Conclusion[20]

It would be sad if blind allegiance to physicalism led to the acceptance of otherwise unmotivated posits of brute absolute necessity between the physical and consciousness. This is especially so in light of the recent resurgence of interest in alternatives to physicalism, such as panpsychism (see Rosenberg (2004), Strawson (2006)) and radical emergentism (e.g. O'Connor and Wong (2005); O'Connor (1994), Silberstein and McGeever (1999)).

In fact, there is a relatively small range of possible responses to the dire situation we find ourselves in with respect to the problem of consciousness which it is worth at least briefly outlining. Readers can position themselves in this spectrum of options.

I label the options 'watchful waiting', 'embrace emergence', 'favour fundamentality' and 'modify metaphysics'.

The first option is the most conservative and perhaps would be the most widely accepted. It requires one to insist that consciousness is a standard conservatively emergent phenomenon that will eventually be fitted into the physicalist picture of the world in the standard way. On this view, it is far too early to junk the scientific view of the world. If the problem persists and grows worse more drastic steps can be taken but the history of science suggests that eventually the problem of consciousness will resolve itself. But by now it is beginning to seem that the general problem of consciousness, the problem of how it is that the physical world embodies or generates phenomenal consciousness, is highly resistant to standard modes of understanding. This issue has attracted sustained philosophical attention for the last fifty years or so[21], under various

[20]This section includes material adapted from the final chapter of Seager (2012b).

[21]Which is not to say that it was not noticed in one way or another long ago; for an historical survey of the problem see Seager (2007).

labels, most recently that of the problem of the 'explanatory gap' (Levine 1983, 2001) and the 'hard problem' (Chalmers 1996, 1997). An enormous literature has ensued but there is no sign of any consensus on a solution, on the exact specification of the problem or even whether it is not some sort of philosophically induced pseudo-problem. But one clear response in the face of epistemic opacity is to plead that our ignorance is excusable and/or explicable. The most straightforward plea is simply that more time is necessary for the study of the brain mechanisms responsible for consciousness, as well as for the advancement of various requisite attendant disciplines—whatever these might be. The depth of our ignorance is perhaps indicated by the fact that candidate disciplines range all the way from physics to philosophy. Moreover, as noted above, a large number of avowed physicalists accept that it is unlikely that consciousness will ever be integrated into the scientific picture in the standard way.

One can follow a more radical path while remaining within the fold of watchful waiting. This path accepts that consciousness will never be successfully integrated into the scientific picture of the world. But this is not because of a problem with consciousness but rather because of an irredeemable intellectual limitation of ourselves. It could be that we suffer from a kind of essential ignorance about the relation between matter and consciousness. As discussed above, this position has been famously defended in McGinn (1989). The obvious weakness of McGinn's approach is that it elevates the scientific picture and conservative emergence to a non-negotiable assumption about the nature of the world.

Yet a third way to adopt watchful waiting, and one that appears to avoid McGinn's pessimistic conclusion, is to predict that science itself will change in a way that will make clear how consciousness is a conservative emergent. There can be no doubt that the explanatory structures within science have changed, sometimes radically, over the course of its development. A salient example is the transmutation in the doctrine of mechanism brought about by the acceptance of Newton's theory of gravitation in the 17^{th} century. The introduction of a non-contact force which acted instantaneously over any distance was regarded with extreme skepticism. Newton could never bring himself to believe in it and wrote in a famous passage: 'that one body may act upon another at a distance through a vacuum, without the mediation of anything else, by and through which their action and force may be conveyed from one to another, is to me so great an absurdity that I believe no man who has in philosophical matters a competent faculty of thinking can ever fall into it' (Janiak 2004, p. 102). Newton's intuition was sound, as the revolution wrought by Einstein eliminated the absurdity.

Thus one might think that the problem of consciousness awaits some new, revolutionary development in science which will unveil the *unmysterious* mechanisms of conservative emergence. Perhaps this is what Noam Chomsky intends to suggest with this analogy:

> Suppose that a nineteenth century philosopher had insisted that 'chemical accounts of molecules, interactions, properties of elements, states of matter, etc. must in the end be continuous with, and harmonious with, the natural sciences,' meaning physics as then understood. They were not, because the physics of the day was inadequate. By the 1930s, physics had radically changed, and the accounts (themselves modified) were 'continuous' and 'harmonious' with the new quantum physics. (Chomsky 2000, p. 82)

Perhaps some similarly radical transformation in physics will allow for consciousness

to take its place as a standard benign emergent. Although Chomsky is in the end noncommittal about this with regard to consciousness, and the mind in general, he does note that '[c]ommonly the "fundamental" science has to undergo radical revision for unification to proceed' (p. 106).

Of course, the greater the hypothetical revolution in science needed to integrate consciousness, the less obvious the resulting edifice counts as the exhibition of standard conservative emergence. Once we admit that a scientific revolution is needed to solve the problem of consciousness, it seems unlikely that the resulting viewpoint will count as a victory for watchful waiting[22].

If watchful waiting is rejected, one can take the bull by the horns and embrace emergence, by which I mean embrace radical emergence. I've argued above that radical emergence requires a genuine distinction between the laws of nature in general and the purely or basic physical laws. If radical emergence is true, a system's behaviour is not determined simply by the physical laws plus the initial conditions. Laws or principles of emergence are also required.

A useful thought experimental model for this stems from computer simulations of physical systems. We already have a number of complex simulations that reveal how complex behaviour emerges from purely physical processes, ranging from quantum chromodynamical calculations of the properties of composite entities (see e.g. Dürr *et al.* (2008)), to large scale climate simulation (extensive discussion can be found in Peixoto and Ort (1992)), to ultra-scale simulations of galactic clusters (Guedes *et al.* (2011)). But the thought experiment asks us to abstract away from the huge computational burdens imposed by vast numbers of components engaging in mind boggingly complex interactions. We are to imagine as perfect as you wish simulation of as extensive as required physical systems; call this the superduper computer simulation thought experiment.

Radical emergentists deny that the superduper simulation would perfectly duplicate any real world system that possesses radical emergence, no matter how well the simulation mimics the lower level. The novel emergent features will have their own causal efficacy which will 'interfere' with the low level processes, forcing them to drift away from the simulated version whose behaviour is, by design, entirely governed by low level processes and low level processes only.

Therefore, radical emergentism maintains there are possible worlds that differ in their 'laws of emergence' without supposing that there is any difference in the subvenient level laws governing these worlds. In terms of our formula [E] this is expressed entirely in the grade of the necessity operator so that two natural forms appear, one using full logical necessity and the other which features only nomological necessity. The former, corresponding to conservative emergence, entails that there are absolutely no possible worlds in which a set of constituents stand in some operative subvening relation but fail to have the associated supervening high level property. The latter, corresponding to radical emergence, allows for worlds in which the various laws of emergence are different, or absent altogether. The emergentist who believes in only conservative emergence thinks the actual world is in the class of worlds that lack any laws of emergence—they

[22]I also am inclined to believe that conservative emergents have a derivative status both in terms of their ontology and their efficacy. Roughly speaking, they are mind-dependent (or observer dependent) epiphenomena. If so, watchful waiting will never succeed given that consciousness is not observer dependent in the relevant sense and it is an efficacious feature of the world. I try to argue for this position in Seager (2012b).

are completely unnecessary. The radical emergentist thinks that a world lacking laws of emergence will not support all the high level features we actually find, notably including consciousness.

Radical emergence seems to be a coherent doctrine (see McLaughlin 1992), although there have been attempts to show otherwise. For example, Thomas Nagel's (1979) argument in favour of panpsychism depends on the denial of the possibility of radical emergence. Nagel's argument is very succinct. The first premise is that 'the properties of a complex system must *derive* from the properties of its constituents, plus they way they are combined' (Nagel 1979, p. 185, my emphasis). All emergentists can agree with this (modulo caveats about the interpretation of 'constituent').

The second premise is that 'true causes *do* necessitate their effects: they make them happen or make them the case' (p. 186). Here Nagel has to assume that the kind of necessitation involved is full logical necessity. But this is highly implausible. What can cause what depends on the laws of nature and various contingent factors. The value of the fine structure constant in in quantum electrodynamics does not appear to be necessitated by any (other) law of nature and it has a powerful influence on the causal powers of atomic structure and hence chemical kinds. As noted by Barrow and Tipler 'it transpires that the gross properties of all atomic and molecular systems are controlled by only two dimensionless *physical* parameters—the fine structure constant... and the electron to proton mass ratio' (1988, pp. 295-6). It seems pretty obviously true that, one, there are causal processes in the actual world that involve these parameters and, two, there are other possible worlds where the causal processes are different because of variation in the values of these parameters.

Nagel is right that once we fix the laws of nature and the state of the world then it is a matter of pure logical necessity what is going to happen and what high level features will appear (if the laws are intrinsically indeterministic then the *range* of outcomes is logically determined). But this does not show that all emergence is conservative emergence unless we assume that the laws of *physics* exhaust the fundamental laws of nature[23]. Since the radical emergentist explicitly denies this Nagel is simply begging the question against this form of emergence. And, in principle, there does not seem to be anything incoherent in the idea that there are irreducible laws of emergence that go beyond the laws of fundamental physics.

I think the main problem facing radical emergence is that it prejudges the possibility of a complete physics. By 'complete' I mean a physics that could generate, in principle, the superduper simulation discussed above. It seems that current physics is on a path towards such a complete theory, and it would be presumptuous to reject this possibility.

Yet another difficulty facing radical emergence is that it appears as an *ad hoc* hypothesis invoked solely to explain the presence of consciousness in the world. This is not how the progenitors of emergentism saw things. Emergentists from Mill onwards saw radical emergence as a pervasive feature of the world operating at the very first level, so to speak, above physics. To the extent that it is now very difficult to argue that chemistry or biology exemplify radical emergence, the only phenomenon which

[23]We have to add the caveat about *fundamental* laws since laws of nature themselves can be conservatively emergent in the usual sense that they are determined by the fundamental laws. For example, the Weidemann-Franz law which states a positive correlative relationship between thermal and electrical conductivity of a metal is an emergent or derived law, which depends on the fact that both heat and electricity conduction depend on the presence of free electrons in metals. The radical emergentist of course posits the existence of primitive, irreducible laws of emergence.

remains a viable candidate for emergence is consciousness. But why should the world have waited so long to exploit its power of radical emergence? Consciousness, with its subjective, first-person aspect, is special. However, this specialness does not in any way suggest that radical emergence is the unique or best way to account for consciousness. It seems that by its nature, radical emergence should appear throughout nature but so far as we can tell it appears nowhere except in an ever shrinking domain where we feel at a total explanatory loss. This makes radical emergence look like a dodge invoked only to plug a gap in understanding with no independent evidence for it whatsoever.

Radical emergence is not the only 'radical' response to the problem of consciousness. Another approach is to accept that consciousness is in some way a fundamental, irreducible feature of the world. This is a venerable approach which has been re-invigorated by recent work, most especially new investigations of panpsychism (see e.g. Blamauer (2011); Skrbina (2009); Brüntrup and Jaskolla (2014)). Double aspect theories and even neutral monist accounts can be considered to fall under the Favour Fundamentality option (the latter since it regards consciousness as equally fundamental as the physical).

In my view, there are many advantages to this approach. The irreducibility of consciousness to the physical is automatically explained. The specialness of consciousness is recognized. At the same time, nothing prevents the possibility of physics generating a complete account of nature at the physical level in which all physical phenomena have physical causes.

One way to see how this latter advantage could be realized is via a Russellian form of panpsychism. Theories of this kind see phenomenal characteristics as the intrinsic nature which, so to speak, rests under and supports the purely structural or relational features which are the sole subject matter of physical science (see Chalmers (forthcominga), Seager (2006)).

Another potential advantage of the panpsychist approach is that it presents what is in some way a minimal alteration in the scientific picture of the world. That is, panpsychism can allow that the physical picture can be complete within itself and it can be formulated so that consciousness is a conservative emergent dependent upon the standard emergence of more complex forms of physical structures (see Seager (2010, 2012a)).

On the other hand, many find panpsychism deeply and viscerally implausible. This seems to be an intuitive reaction which sometimes shortcircuits reflection on the topic (see e.g. Searle (1997): panpsychism is absurd, or McGinn (1999): panpsychism is ludicrous). However, panpsychism also faces a number of serious objections that go beyond the 'incredulous stare'. Perhaps the most potent of these is the complaint that there is no way for the panpsychist to provide a scheme by which the putative mental aspect or component of the microfeatures of the world can combine to generate the sorts of experiences such as conscious beings such as ourselves enjoy (see Seager (1995), Chalmers (forthcomingb)). The panpsychist here seems to face a severe problem: the combination function that has to be posited looks suspiciously like a relation of radical emergence, but if radical emergence is acceptable why not take the apparently simpler route of letting consciousness radically emerge from the physical?

Finally, one can modify one's metaphysics. In some ways, this is the most radical reaction to the problem of consciousness. Crudely speaking, this option requires one to give up a kind of ultra-strong scientific realism, or as it might be more pejoratively called, scientism which is not unfamiliar: the view that science is the sole source of

knowledge about the ultimate nature of reality. That science, and especially the science of physics, has this role is widely held both among scientists, philosophers and our culture more generally (as evidenced by the sort of pronouncements given in note 10 above). To quote just one current statement of the position, 'Any new metaphysical claim that is to be taken seriously should be motivated by, and only by, the service it would perform, if true, in showing how two or more specific scientific hypotheses jointly explain more than the sum of what is explained by the two hypotheses taken separately, where a "scientific hypothesis" is understood as an hypothesis that is taken seriously by institutionally bona fide current science' (Ladyman *et al.* (2007), p. 60). Nonetheless, there are more or less well developed positions in philosophy of science that see science in a different light. I am thinking here particularly of Bas van Frassen's constructive empiricism (see van Fraassen (1980); van Fraassen (2002)), Nancy Cartwright's views on the laws of nature (see Cartwright (1983, 1999)) and John Dupré's account of the disunity of the world (see Dupré (1993)).

I would characterize this position as one in which the world is taken to be as it is given in our ordinary experience. Ultimate reality is not to be sought in anything transcending experience. Science can help us understand the structure and processes in this world, but this help is in the form of a large set of diverse models applicable to observation and experimentation. Scientific explanations are model based and it is presumed that the models are but imperfect replicas of small portions of reality which are tailored to the specific job of prediction they are designed to aid, nor do these models point to a comprehensive view of the world which is an accurate picture of ultimate reality.

In such a reoriented vision of science, consciousness takes its place as part of reality many of whose features can be scientifically explained. We can investigate the neural correlates of consciousness, the physical conditions necessary for consciousness to appear, the evolution of consciousness, etc. But the fact–if it is a fact–that the phenomenal or essentially subjective aspect of consciousness cannot be explained scientifically, nor integrated into some putatively total and comprehensive picture of reality based entirely on fundamental physics, is not a deep worry. Not everything in the world is amenable to the scientific modeling project.

Obviously, to many thinkers, such a radical re-visioning of the nature and role of basic science is completely unacceptable and, of course, there are many strong objections to the anti- or a-realist positions. But I think it does provide an interesting and very different way to tackle the impasse generated by the problem of consciousness.

In conclusion, I've tried to argue for two main propositions. The first is that the burden of proof in this debate rests on the shoulders of the physicalists. This may not have always been so, but the long standing failure to show how consciousness is reductively epistemologically dependent on the physical has by now shifted the burden. The second is that in a curious way the success of physicalism heretofore is in a way 'undercutting'. This history of this success has uniformly proceeded by exhibition of the mechanisms of epistemological dependence. The kind of barrier which consciousness has placed in the path to completion of the physicalist picture of the world is one that flatly blocks such exhibition. Which in turn suggests that there may well be some kind of modal independence between consciousness and the physical world. Either that, or our understanding of the physical world is deeply incomplete at the moment. Thus, at the very least, we should welcome the exploration of a variety of non-physicalist accounts of the metaphysical structure of the world.

William Seager
University of Toronto Scarborough

References

Alexander, Samuel (1920). *Space, Time and Deity*. Macmillan & Co.

Alexander, Samuel (1922). 'Natural Piety'. *The Hibbert Journal*, 20 (4): pp. 609–27.

Armstrong, David (1968). *A Materialist Theory of the Mind*. London: Routledge and Kegan Paul.

Banegas, Diane *et al.* (2007). 'At Discovery's Horizon: Report of the Task Force on US LHC Communication'. Tech. rep., National Science Foundation. URL `http://www.lhcus.org/communication/documents/US_ILC_Report_101807_v3.pdf`.

Barrow, John and Frank Tipler (1988). *The Anthropic Cosmological Principle*. Oxford: Oxford University Press.

Belot, Gordon and J. Earman (1997). 'Chaos out of Order: Quantum Mechanics, the Corrspondence Principle and Chaos'. *Studies in History and Philosophy of Modern Physics*, 28: pp. 147–182.

Blamauer, Michael (ed.) (2011). *The Mental as Fundamental: New Perspectives on Panpsychism*. Frankfurt: Ontos Verlag.

Block, Ned and Robert Stalnaker (1999). 'Conceptual analysis, dualism, and the explanatory gap'. *Philosophical Review*, 108: pp. 1–46.

Broad, C. D. (1925). *Mind and Its Place in Nature*. London: Routledge and Kegan Paul.

Brock, William (1993). *The Norton History of Chemistry*. New York: W. H. Norton & Co.

Brüntrup, G. and L. Jaskolla (eds.) (2014). *Panpsychism*. Oxford: Oxford University Press.

Cahn, Robert N. (1996). 'The eighteen arbitrary parameters of the standard model in your everyday life'. *Rev. Mod. Phys.*, 68: pp. 951–59. URL `http://link.aps.org/doi/10.1103/RevModPhys.68.951`.

Carroll, Sean (2010). 'The laws underlying the physics of everyday life are completely understood'. *Cosmic Variance*, September 23. URL http://blogs.discovermagazine.com/cosmicvariance/2010/09/23/the-laws-underlying-the-physics-of-everyday-life-are-completely-understood/.

Cartwright, Nancy (1983). *How the Laws of Physics Lie*. Oxford: Oxford University Press (Clarendon).

Cartwright, Nancy (1999). *The Dappled World*. Cambridge: Cambridge University Press.

Chalmers, David (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.

Chalmers, David (1997). 'Facing Up to the Problem of Consciousness'. In Jonathan Shear (ed.), *Explaining Consciousness*, pp. 9–32. Cambridge, MA: MIT Press.

Chalmers, David (forthcominga). 'Panpsychism and Panprotopsychism'. In T. Alter and Y. Nagasawa (eds.), *Consciousness in the Physical World: Essays on Russellian Monism*. Oxford: Oxford University Press.

Chalmers, David (forthcomingb). 'The Combination Problem for Panpsychism'. In G. Brüntrup and L. Jaskolla (eds.), *Panpsychism*. Oxford: Oxford University Press.

Chomsky, Noam (2000). *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press.

Dennett, Daniel (1991). *Consciousness Explained*. Boston: Little, Brown & Co.

Descartes, René (1641/1985). 'Meditations of First Philosophy'. In J. Cottingham, R. Stoothoff and D. Murdoch (eds.), *The Philosophical Writings of Descartes, Vol. 2*, pp. 1–62. Cambridge: Cambridge University Press. (J. Cottingham, R. Stoothoff, D. Murdoch, trans.).

Dobzhansky, Christian (1973). 'Nothing in Biology Makes Sense Except in the Light of Evolution'. *American Biology Teacher*, 35: pp. 125–29.

Dupré, John (1993). *The Disorder of Things*. Cambridge, MA: Harvard University Press.

Dürr, S. *et al.* (2008). 'Ab Initio Determination of Light Hadron Masses'. *Science*, 322 (5905): pp. 1224–27.

Flanagan, Owen (1991). *The Science of the Mind*. Cambridge, MA: MIT Press, 2nd ed.

Friesner, Richard (2005). '*Ab initio* quantum chemistry: Methodology and applications'. *Proceedings of the National Academy of Sciences*, 102 (19): pp. 6648–53.

Goodale, Melvyn and David Milner (2004). *Sight Unseen*. Oxford: Oxford University Press.

Guedes, Javiera, Simone Callegari *et al.* (2011). 'Forming realistic late-type spirals in a ΛCDM universe: the Eris simulation'. *The Astrophysical Journal*, 742 (2).

Hardin, C. L. (1988). *Color for Philosophers: Unweaving the Rainbow*. Indianapolis: Hackett.

Henson, Shandelle, R. Constantino *et al.* (2001). 'Lattice Effects Observed in Chaotic Dynamics of Experimental Populations'. *Science*, 294: pp. 602–605.

Jackson, Frank (1982). 'Epiphenomenal qualia'. *Philosophical Quarterly*, 32: pp. 127–36.

Janiak, Andrew (ed.) (2004). *Newton: Philosophical Writings*. Cambridge: Cambridge University Press.

Kripke, Saul (1980). *Naming and Necessity*. Cambridge: Cambridge University Press.

Ladyman, James, Don Ross *et al.* (2007). *Everything Must Go: Metaphysics Naturalized*. Oxford: Oxford University Press.

Levine, Joseph (1983). 'Materialism and Qualia: The Explanatory Gap'. *Pacific Philosophical Quarterly*, 64: pp. 354–61.

Levine, Joseph (2001). *Purple Haze: The Puzzle of Consciousness*. Oxford: Oxford University Press.

Loar, Brian (1990). 'Phenomenal States'. In J. Tomberlin (ed.), *Philosophical Perspectives*, vol. 4, pp. 81–108.

McGinn, Colin (1989). 'Can we solve the mind-body problem?' *Mind*, 98 (391): pp. 349–366.

McGinn, Colin (1993). *Problems in Philosophy: The Limits of Inquiry*. Oxford: Blackwell.

McGinn, Colin (1999). *The Mysterious Flame: Conscious Minds in a Material World*. New York: Basic Books.

McLaughlin, Brian (1992). 'The Rise and Fall of British Emergentism'. In A. Beckermann, H. Flohr and J. Kim (eds.), *Emergence or Reduction*, pp. 49–93. Berlin: De Gruyter.

Meehl, Paul (1966). 'The Compleat Autocerebroscopist: A Thought-Experiment on Professor Feigl's Mind-Body Identity Thesis'. In P. Feyerabend and G. Maxwell (eds.), *Mind, Matter and Method: Essays in Philosophy and Science in Honor of Herbert Feigl*, pp. 103–180. Minneapolis: University of Minnesota Press.

Montero, Barbara (2005). 'What is the Physical?' In *The Oxford Handbook of Philosophy of Mind*. Oxford University Press.

Morgan, Conwy Lloyd (1923). *Emergent Evolution*. London: Williams and Norgate.

Nagel, Thomas (1974). 'What Is It Like to be a Bat?' *Philosophical Review*, 83 (4): pp. 435–50. (This article is reprinted in many places, notably in Nagel's *Mortal Questions*, Cambridge: Cambridge University Press, 1979.).

Nagel, Thomas (1979). 'Panpsychism'. In *Mortal Questions*, pp. 181–95. Cambridge: Cambridge University Press. (Reprinted in D. Clarke *Panpsychism: Past and Recent Selected Readings*, Albany: SUNY Press, 2004.).

Nagel, Thomas (1986). *The View from Nowhere*. Oxford: Oxford University Press.

Nagel, Thomas (2012). *Mind and Cosmos*. Oxford: Oxford University Press.

Ney, Alyssa (2008). 'Physicalism as an attitude'. *Philosophical Studies*, 138 (1): pp. 1–15.

Nishimoto, Shinji, An T. Vu *et al.* (2011). 'Reconstructing Visual Experiences from Brain Activity Evoked by Natural Movies'. *Current Biology*, 21 (9): pp. 1641–6.

O'Connor, Timothy (1994). 'Emergent Properties'. *American Philosophical Quarterly*, 31: pp. 91–104.

O'Connor, Timothy and Hong Yu Wong (2005). 'The Metaphysics of Emergence'. *Noûs*, 39: pp. 658–678.

Owen, Adrian M. (2008). 'Functional neuroimaging of the vegetative state'. *Nature Reviews Neuroscience*, 9 (3): pp. 235–43.

Papineau, David (2006). 'Phenomenal and Perceptual Concepts'. In T. Alter and S. Walter (eds.), *Phenomenal Concepts and Phenomenal Knowledge*. Oxford: Oxford University Press.

Peixoto, José and A. Ort (1992). *Physics of Climate*. New York: American Institute of Physics.

Rey, Georges (1993). 'Sensational Sentences'. In M. Davies and G. Humphreys (eds.), *Consciousness*, pp. 240–57. Oxford: Blackwell.

Rosenberg, Gregg (2004). *A Place For Consciousness: Probing the Deep Structure of the Natural World*. Oxford: Oxford University Press.

Seager, William (1995). 'Consciousness, Information and Panpsychism'. *Journal of Consciousness Studies*, 2 (3): pp. 272–88. (Reprinted in J. Shear (ed.) *Explaining Consciousness*, Cambridge, MA: MIT Press, 1997.).

Seager, William (2006). 'The 'intrinsic nature' argument for panpsychism'. *Journal of Consciousness Studies*, 13 (10-11): pp. 129–45. (Reprinted in A. Freeman (ed.) *Consciousness and Its Place in Nature*, Exeter: Imprint Academic, 2006.).

Seager, William (2007). 'A Brief History of the Philosophical Problem of Consciousness'. In P. Zelazo, M. Moscovitch and Evan Thompson (eds.), *The Cambridge Handbook of Consciousness*, pp. 9–34. Cambridge: Cambridge University Press.

Seager, William (2010). 'Panpsychism, Aggregation and Combinatorial Infusion'. *Mind and Matter*, 8 (2): pp. 167–84.

Seager, William (2012a). 'Emergentist Panpsychism'. *Journal of Consciousness Studies*, 19 (9-10): pp. 19–39.

Seager, William (2012b). *Natural Fabrications: Science, Emergence and Consciousness*. Berlin: Springer Verlag (Frontiers Series).

Searle, John (1997). 'Consciousness and the Philosophers'. *New York Review of Books*, 44 (4): pp. 43–50.

Sellars, Wilfrid (1963). 'Philosophy and the Scientific Image of Man'. In *Science, Perception and Reality*, pp. 1–40. London: Routledge and Kegan Paul.

Siewert, Charles (1998). *The Significance of Consciousness*. Princeton, NJ: Princeton University Press.

Silberstein, Michael and J. McGeever (1999). 'The Search for Ontological Emergence'. *Philosophical Quarterly*, 49: pp. 182–200.

Sklar, Lawrence (1993). *Physics and Chance: Philosophical Issues in the Foundations of Statistical Mechanics*. Cambridge: Cambridge University Press.

Skrbina, David (ed.) (2009). *Mind That Abides: Panpsychism in the new millennium*. Amsterdam: John Benjamins.

Stoljar, Daniel (2001). 'Two Conceptions of the Physical'. *Philosophy and Phenomenological Research*, 62: pp. 253–81.

Strawson, Galen (2006). 'Realistic Monism: Why Physicalism Entails Panpsychism'. *Journal of Consciousness Studies*, 13 (10-11): pp. 3–31. (Reprinted in A. Freeman (ed.) *Consciousness and Its Place in Nature*, Exeter: Imprint Academic, 2006.).

van Fraassen, Bas (1980). *The Scientific Image*. Oxford: Oxford University Press (Clarendon).

van Fraassen, Bas (2002). *The Empirical Stance*. New Haven: Yale University Press.

White, Stephen (1986). 'Curse of the Qualia'. *Synthese*, 68 (2): pp. 333–68.

White, Stephen (2010). 'The Property Dualism Argument'. In R. Koons and G. Bealer (eds.), *The Waning of Materialism*, pp. 89–114. Oxford: Oxford University Press.

Wilson, Jessica M. (2006). 'On Characterizing the Physical'. *Philosophical Studies*, 131 (1): pp. 61–99.