1 Consciousness and physicalism

CHAPTER 1 • PREVIEW

Physicalism is the view that everything that exists is ultimately physical. It is the dominant metaphysics of nature at present despite facing a number of formidable challenges. Here I examine the reasons we have for believing in physicalism. It turns out that the undeniable success of physicalism may in fact undercut the claim that physicalism deserves wholesale 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 towards 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.

1.1 WHAT IS PHYSICALISM?

The problem of consciousness is special. We have intimate knowledge of it. Writers and thinkers have spent vast efforts in outlining its structure, capabilities and forms, in fiction, science and philosophy. Spectacular advances have been made in recent times in our ability

to correlate brain processes with various kinds of conscious experiences. We can now tell, with high reliability, whether someone is conscious or not by using MRI brain imaging, even if the subject has been diagnosed as being in a vegetative state and gives no behavioural signs of anything but profound coma (see Owen 2008). Of course, we do not have a detailed account of how or where consciousness resides in the brain nor any such account of exactly what systems within the brain consciousness particularly or specifically depends upon.[1] But nor do we have such accounts of any mental functions as yet, be they conscious or unconscious. The brain is vastly complicated and our investigations can still only be called preliminary but they are proceeding apace. We are truly on the way to finding out everything about consciousness that science can tell us.

Yet the problem of consciousness is special. It is the spectre haunting the scientific view of the world. The fear is that science cannot tell us everything there is to know about consciousness. This would be in conflict with an otherwise very attractive view of the world whose success has been spectacularly extensive and cumulative over the last half millennium.

This book is not directly about this view – called 'materialism' or, more accurately, 'physicalism', which is a philosophical interpretation of the success of physical science. The literature on physicalism: its nature, commitments, strengths and problems is staggeringly vast. In this book I do not aim to add anything definitive to this particular body of work in the way of settling any of these issues. But physicalism will loom large in the background since virtually all the philosophical theories of consciousness to be considered were formulated with at least one eye on the question of how to integrate consciousness into a physicalist account of the world. Some others to be considered result from a perceived failure of physicalism but even they must respect the scope and power of the physicalist world view.

It is important, then, to begin with an overview of the metaphysical problem of consciousness which puts in place the nature and goals of physicalism and highlights how consciousness generates a unique problem for it. This is a daunting task, but since we require here only a bird's eye view, and since that will be more than sufficient to show why the problem of consciousness is special, the task is approachable. Beyond characterizing physicalism, the main issue to consider is why physicalism is favoured amongst so many and to ask whether its favoured status is really justified.

Physicalism is a monistic metaphysics: it claims that there is only one basic kind of reality which is physical in nature. However, the nature of the physical is another very vexed philosophical issue (see e.g. Montero 2009, Strawson 2006, Stoljar 2001, Howell 2013, pt. 1, Wilson 2006). 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 naive intuition suggests 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 (and ostensibly) 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).

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, non-social, non-economic, 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.

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 2012 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 do not yet see such interconnections or anchor points is a sore point, like a nagging splinter, but there remain very few such problematic areas.

It is easy to find working physicists who espouse this grand view. For example, the well-known string theorist and science popularizer Brian Greene believes that any 'physical system is completely determined by the arrangement of its particles' (2011, p. 38). This is not to be read as a mere definition of 'physical system' which leaves open whether there are non-physical systems lurking in reality, for Greene explicitly avows what he calls the 'reductionist view' which he notes is 'common among physicists'. He elaborates that 'the position that makes the most sense to me is that one's physical and mental characteristics are nothing but a manifestation of how the particles in one's body are arranged' (2011, p. 39). Even physicists such as Philip Anderson, famous for his anti-reductionism, does not dispute the sort of reductionism outlined by Greene. In his article, 'More is Different', Anderson begins with the clarification that:

> The reductionist hypothesis may still be a topic for controversy among philosophers, but among the great majority of active scientists I think it is accepted without question. The workings of our minds and bodies, and of all the animate or inanimate matter of which we have any detailed knowledge, are assumed to be controlled by the same set of fundamental laws . . .
>
> (1972, p. 393)

In the abstract, physicalism thus demands that there be a dependence (or determination) 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: logically and ontologically.

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. The basic form of this relation is that of logical supervenience.

The concept of supervenience has become a complex topic in philosophy over the last 40 years or so, since it was reintroduced into the philosophers' arsenal by Donald Davidson (1970; for an overview see McLaughlin and Bennett 2008). For our purposes, it is enough to understand logical supervenience as asserting that for any non-fundamental property, there are fundamental physical conditions which absolutely guarantee its instantiation. If you duplicate these physical conditions you will get a new instance of the property in question. And if you want to alter the distribution of the property in question, then you must make some change in the relevant fundamental physical conditions. Note that a claim of supervenience does not entail any claims about explicability. It is, so to speak, a purely metaphysical relation which asserts the determination of the non-fundamental by the fundamental but does not venture to say how the determination comes about.

Logical supervenience is 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'. 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. It is important to emphasize that 'small' does not necessarily mean 'particle'. The relation between what our fundamental physics tells us about the world – that it is composed of a myriad of quantum fields – and a notion that the world is made up of very small pieces of matter is very far from straightforward. As David Wallace has written: 'the popular impression of particle physics as about the behaviour of lots of little point particles whizzing about bears about as much relation to real particle physics as the earth/air/fire/water theory of matter bears to the Periodic Table' (2013, p. 222; see also Fraser 2008 for a technical argument that particles cannot be considered as fundamental in quantum field theory). 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 remain outside of our reach for all save the very simplest cases.[2] 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 property, *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*:

$$[\mathrm{E}]Fa \rightarrow (\exists\Gamma)(\exists z_1, z_2, \ldots, z_n)(Cz_1, z_2, \ldots, z_n, a \wedge \Gamma z_1, z_2, \ldots, z_n \wedge \Box_l(\forall x_1, x_2, \ldots, x_n, y)(Cz_1, z_2, \ldots, z_n y \wedge \Gamma z_1, z_2, \ldots, z_n \rightarrow Fy)).$$

What this formula says is that if an object, *a*, has a property, *F*, then there is a relation which holds amongst its constituents (whatever they may be) such that any such system of constituents related by that relation will necessarily have the property, *F*.

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 suggests that the properties of things are not always 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[3] inasmuch as an object comes to possess a property which its constituents lack. Physicalism will have to embrace some form of emergence.

There are various kinds of emergence but a broad division will suffice for our purposes. Conservative emergence, of which [E] is a partial characterization, essentially claims that emergent properties are simply new ways to describe complex situations. Such new descriptions cannot be directly predicted from underlying theory. Nothing in atmospheric dynamics, for example, predicts the concept 'thunderstorm'. But if one was given a new descriptive concept ('thunderstorm') and a simulation of the world based solely on fundamental laws (a simulation of the atmosphere on a hot, dry afternoon in Alberta say), one would see that complexes in the simulation deserved to be described by the new concept (things acting just like thunderstorms would appear spontaneously in the simulation). Radical emergence goes further, asserting that the emergent properties make a real difference to the workings of the world. Radical emergentism claims that the simulation based only on fundamental physical law would simply fail to simulate the world accurately. If thunderstorms were radically emergent, the atmospheric simulation, no matter how perfect, would go on and on, but never generate anything like a thunderstorm.[4]

As noted, the properties deployed in fundamental physics are but a tiny fraction of the properties the world exemplifies. But there need be nothing mysterious, mystical or transcendent about the conservative emergence vouchsafed by [E]. No physicalist should be worried about this kind of emergence and in fact they should welcome its inclusion in the physicalist world view.

The strength of the necessity operator, $\Box_l$, is crucial here. It must possess the maximal modal strength of logical necessity (hence the mnemonic 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 latter view, that emergence should be understood in terms of a supervenience relation defined via nomological necessity rather than strict logical necessity is perfectly respectable and not unfamiliar. In essence, it was the view held by the so-called British emergentists, notably Samuel Alexander (1920), Conwy Lloyd Morgan (1923) and C. D. Broad (1925) (for a general discussion see McLaughlin 1992). 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 up the physical laws and the arrangement of the fundamental physical entities in the world, there was nothing left to do. They would follow of necessity as a metaphysical 'free lunch'. If some 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.[7] 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' (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 if and only if 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.

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, statistical mechanics 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 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 with the concepts characteristic of the reduced 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 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 relatively puny human scientists, 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.

1.2 EVIDENCE FOR PHYSICALISM

If one steps back from any pre-existing philosophical 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 many 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' 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' (1963b, 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 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 . . .
>
> (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 labelled 'the optimistic induction'.[8] It is based simply on the fact that the scientific enterprise has, for over 400 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 nineteenth 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.[9] 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 physicalist picture yet more seriously. By embracing the idea of literally 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 (see Papineau 2000 for an influential general development of this line of argument).

However, for our purposes we need not dwell on this long history of success. What is important to note is that throughout this 400-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 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.[10]

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 incomplete autonomy of the higher-level and the occasional breakdown of its laws is taken as a sign that it is the fundamental level that is really driving the system. Intrusions suggest, but do not prove, that all the 'control knobs' for higher-level phenomena reside at the level of the physically fundamental.

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 under-labourer to science. Overall, given that the foundational science of physics is taken to provide a pretty good account of the most 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. For example, such an idea is exploited by science lobbyists, as illustrated by a remark from the United States LHC (Large Hydron Collider) 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). Thus 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.

1.3 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 2012, 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 signalling future abilities in the realm of 'mind reading' the social and ethical implications of which deserve careful reflection. 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'[11] (1974), Frank Jackson's 'knowledge argument' (1982) and Descartes'/Saul Kripke's/David Chalmers' 'modal-conceivability argument' (see Descartes 1985c, Kripke 1980, Chalmers 1996).

Like all great philosophical arguments, the anti-physicalist arguments can be expressed in simple and intuitively compelling forms.

Nagel's argument can be summarized as follows:

(1) States of consciousness have a subjective feature which accounts for why there is something it is like to be in them.
(2) These subjective features are not present in fundamental physics.
(3) None of the mechanisms of conservative emergence can generate subjective features from the objective features present in fundamental physics.
(4) Therefore, consciousness cannot be integrated into the physicalist view of the world.

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] might be true' (1974, p. 177). Nagel is officially agnostic

about the truth of physicalism, or even leans towards accepting it, but takes it for granted that in the absence of 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).

An abstract summary of Jackson's knowledge argument might go like this:

(1) It is possible to acquire complete knowledge of the physical facts.
(2) Even such complete physical knowledge would not provide knowledge of subjective features of states of consciousness which accounts for why there is something it is like to be in them.
(3) Therefore, there are facts (objects of knowledge) that go beyond the physical facts.

For intuitive vivacity Jackson casts the argument in the form of a puzzle about a particular super-scientist, Mary, who has never had any colour experiences. Her compendious knowledge of the physics of light and vision will not suffice to give her knowledge of what it is like to experience red and she will learn something new the first time she sees a ripe tomato. To most people, this seems obviously true and the effort to undercut this intuition takes some serious philosophical footwork.

One can regard Jackson's argument about the physically omniscient neuroscientist from the epistemological dependence 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 (but note it was not ludicrous in 1910). 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 the 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.[12]

The final argument of the mighty triumvirate is the conceivability argument which goes back to René Descartes' sixth meditation. It has been updated most recently by Kripke and Chalmers. In bare outline, it may be summarized thus:

(1) It is conceivable that consciousness could vary independently of physical state.
(2) Thus it is possible that consciousness could vary independently of physical state.
(3) Therefore, consciousness is not ontologically dependent on the physical.

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.

These arguments can be viewed as travelling 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 and not, in the first instance, epistemological dependence. The classic arguments are in themselves hardly 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 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 trio of arguments show 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. Without such a relation, the link between the physical world and consciousness threatens to involve a measure – the extent of which is unknown – of contingency. Why could not consciousness be distributed slightly differently, or vastly differently as in the zombie scenario,[13] in a world that was physically identical to the actual world? Without exhibition of the relation of epistemological dependence between the physical and the phenomenal, the explanatory gap threatens to turn into an ontological gap.[14] The whole dialectic is in a peculiar position.

1.4 THE BURDEN OF PROOF

Does the long history of physicalist success make physicalism the 'default' position? Although many scientists and philosophers take this stand, it 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 a reductive epistemological dependence relation to the physical.

This is the proposition for which we have historical inductive evidence. The basic 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 catalogue 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 a 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.[15]

Furthermore, there are famous relations of metaphysical independence that can mimic the appearance of dependence, namely 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 which will help clarify and illuminate this claim. What we have established is that the long history of physicalist success inductively implies that:

$Pr(E \mid P)$ is very high,

where E = 'consciousness is reductively epistemologically dependent on the physical' and P = 'physicalism is true'. Using the definition of conditional probability, we can express this in a mathematically equivalent way, as:

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

Since the probability of a conjunction is equal to the probability of the first conjunct multiplied by the probability of the second conjunct conditional upon the first, this latter expression can then be rewritten as:

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

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

To get a handle on this, let us suppose, for the sake of the argument, 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, clarified, deepened and bolstered by the three classic arguments reviewed above.

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 1989a, Pinker 1999).

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 substantial doubts on the truth of physicalism. And consciousness appears to be exactly this kind of roadblock.

1.5 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, consider mysterianism (see e.g. McGinn 1989a, 1999, 1991, 1993).[16] Mysterianism claims that it is our own intrinsic intellectual weakness that prevents us from understanding how consciousness is epistemologically dependent on the physical. It 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 by, for example, some intellectually superior alien scientists.

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* knowledge. 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. 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 two distinct sets 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. 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).

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 demeanour of Clark Kent versus those 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 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 or primitive recognitional capacity which does not depend on appreciation of any feature of experience. In general, the existence of such capacities does not seem particularly hard to accept. The by now familiar phenomenon of blindsight is a striking example. As is well known, certain sorts of brain damage to the visual centres of the brain can create scotomas in 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.[17]

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 the 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 you are are experiencing. 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.[18] It is of course striking that what is missing in these cases is specific sorts of consciousness despite the presence of certain recognitional capacities.

It would be natural to reply to this line of thought that there must be direct recognitional concepts on pain of a vicious regress. The regress threatens if we take it that, say, S recognizes P via a mode of presentation of P, call it Q. How does S recognize Q? We cannot just say this is accomplished via recognition of a further mode of presentation of Q without generating the regress. Brian Loar put it thus:

> Even on the antiphysicalist view, phenomenal concepts are recognitional concepts, and we have 'direct' recognitional concepts of phenomenal qualities . . . it would be absurd

> to insist that the antiphysicalist hold that we conceive of a phenomenal quality of one kind via a phenomenal mode of presentation of a distinct kind.
>
> (1990, p. 228)

If we are forced to allow direct recognitional concepts, then why should the physicalist not have every right to claim that the referent of such concepts is a purely physical state or property? It seems to me that the physicalist can indeed make this claim, but if so the claim will entail that the physical world has phenomenal qualities. That is, this claim will amount to the claim that the physical world (or some of it at any rate, e.g. brains) has 'presentational' properties which constitute aspects of conscious experience. The fact that properties of phenomenal consciousness are directly recognizable does not entail that they are empty of presentational content. It is evident that our acquaintance with phenomenal qualities is not a mere empty presentiment.[19] It may well be that matter has the feature we apprehend via conscious experience as one of its characteristics, albeit a characteristic inaccessible from, as it were, the outside. It might even be a feature of fundamental physical reality, as panpsychists maintain (see Chapters 13 and 14 below for a discussion of panpsychism).

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. Recognitional capacities trigger the application of these concepts and discursive thought over the long span of human cognitive development has elaborated them into a structure which supports a rich but delusive system of beliefs. In terms of what we think consciousness is *within* this system, we are actually no more conscious than rocks.

A 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.
>
> (1991a, p. 98)

I hesitate to ascribe this bald view to Dennett since his writing is often ambiguous between a position that is solely devoted to debunking certain supposedly dubious philosophical notions, such as that of qualia (this issue is discussed at length in Chapters 7 and 8 below), 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 inter-relations of more ordinary physical objects.

The idea that presence is a fictional object seems too wildly implausible to be taken seriously, yet it seems to be the natural upshot of the pure recognitional capacities 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 (see Chalmers 1996, ch. 4 and Chalmers 2010, ch. 6 for extensive criticism of brute necessities). It entails abjuring any prospect of genuinely 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.

It also 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. 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 has an obvious odour 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 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 'catalogue of brutality' is provisional, but the fundamental features of the world, whatever they may be, are brute facts, inexplicable and to be accepted with 'natural piety'.[20] For example, the standard model of physics, our most comprehensive and successful fundamental physical theory contains some eighteen parameters whose values are, theoretically speaking, arbitrary – they must be empirically determined (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 and the aforementioned Cahn 1996) which has led to fascinating cosmological insights.

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 seemingly 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 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.

This line of argument fails here, however, because it neglects the fact that whatever physical state is to be 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.

In this chapter I have 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'. The 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. This, 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.

In what follows we shall see this problem played out in detail. Many if not most philosophical theories of consciousness aim to integrate consciousness into the kind of physicalist picture of the world outlined above. I think it would be fair to say that these theories have the ultimate aim of either showing how consciousness is epistemologically dependent upon the physical or, increasingly more common, how the failure of this dependence can be in some way excused and physicalism maintained despite its absence.

We shall also examine some theories of consciousness that see the failure to successfully demonstrate either of these disjuncts as grounds for exploring more radical, non-physicalist approaches.

But, to begin, let us return to the progenitor of the modern problem of consciousness. René Descartes was the first to see that the physicalist view of the world was deeply at odds

with the phenomenon of consciousness, and I think his insights still shape current debate both about the nature of consciousness and its relation to the physical world.

CHAPTER 1 • SUMMARY

Although physicalism is in a strong antecedent position it has heretofore always provided physical explanations for all the phenomena it has grappled with. Consciousness stands as perhaps the single natural fact on which physicalism cannot get an explanatory grip. Many physicalists agree that in the case of consciousness we will not get the usual physical assimilation. But if so, this would strongly undercut confidence in physicalism whose strength depends upon its explanatory success. There are many ways to address this worry but it is legitimate to wonder if the phenomena of consciousness will lead to the downfall of physicalism.

NOTES

1 Our techniques for detecting conscious states remain very crude, but already brain-based evidence is compelling. For example, it is possible to communicate with some patients misdiagnosed as being in a vegetative state via their deliberate attempts to engage in certain sorts of conscious activities (imagining playing tennis for example). This is a kind of weak mind reading (see Monti *et al*. 2010). Although such results are very far from pinpointing the brain states which are responsible for consciousness, no one but a philosophical sceptic could deny that we are able to detect consciousness via brain imaging in such cases.

2 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, p. 6651)

3 The general topic of emergence is another complicated and much discussed issue in philosophy and science. For an excellent overview, see O'Connor and Wong (2012).

4 Remarkable simulations of the emergence of large scale atmospheric phenomena based upon micro-dynamics already exist. For a spectacular tornado simulation see Orf (2014).

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 illustrate this distinction is the metaphysical but non-logical necessity of the identity of water and H_2O (this sort of a

posteriori necessity was first brought to philosophical prominence by Saul Kripke 1980). 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 sharing our fundamental physical laws and, broadly speaking, conditions where there is H_2O but no water. For a classic discussion of the status of metaphysical necessity see Chalmers (1996), ch. 4.

6 But it is 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.

7 For an extensive discussion of epistemological dependence (or something very much like what I mean here) see Jackson (1998) and the exchange between Block and Stalnaker (1999) and Chalmers and Jackson (2001).

8 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.

9 For more on this story see Brock (1993), ch. 5.

10 For more on the Loschmidt–Boltzmann controversy see Sklar (1993).

11 This evocative way of characterizing consciousness was canonized by Nagel, but there were philosophical precursors. Ludwig Wittgenstein: 'I know what it's like to see red' (1980a, v. 1, § 91 [the work dates back to the mid-1940s]), Brian Farrell: '[we] wonder what it would be like to be one of them' (1950, p. 177), Timothy Sprigge: 'consciousness is . . . what it is or might be like to be a certain object' (1971, p. 168).

12 For a set of discussions of the knowledge argument see Ludlow *et al.* (2004), including an interesting diagnosis of the power of the argument (Hellie 2004) which has broken into popular culture recently in the film *Ex Machina*.

13 Philosophical zombies, unlike the Caribbean or Hollywood varieties, are perfect physical replicas of conscious beings which are nonetheless entirely unconscious. They act exactly like us in every respect, complain about pains, pursue pleasures, but all is darkness within. If zombies are possible then physicalism cannot be true for they show that consciousness can vary independent of the physical. For a seminal discussion of the zombie argument see Kirk (1974); a general overview can be found in Kirk (2012). It is worth noting that Kirk's views on the zombie argument have reversed; his anti-zombie position is argued at length in his (2005).

14 The idea of the explanatory gap and the initial exploration of its significance for the physicalism debate stems from Joseph Levine (see 1983; 2001).

15 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 that there are nomological relations between the structure and processes within my brain and my states of consciousness which preclude such local, intra-world variation.

16 The label of 'mysterianism' was first bestowed on this doctrine by Owen Flanagan (1991), sparked by the name of the now obscure rock band. An extended version presented in popular terms can be found in John Horgan (1999).

17 For a scientific overview of blindsight and associated phenomena see Weiskrantz (1997) and Goodale and Milner (2004). For a detailed investigation of the relevance of various blindsight thought experiments to the problem of consciousness, see Siewert (1998).

18 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.

19 This term figures in Dennett's account of conscious perception, which might be regarded as an instance of the empty direct recognition strategy (see 1978b); an interesting critique which emphasizes the 'failure of fit' between perceptual phenomenology and this sort of account of perceptual experience is McDowell (1994b).

20 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' (1922, p. 609).