Chapter 2 – What is Physicalism?

The first half of this book is spent arguing against physicalism, the view that fundamental reality is wholly physical. Physicalism is an extremely popular position in analytic philosophy of mind, and many of the central debates in this area are between physicalists and their
opponents. However, there is a great irony here: it is not at all clear upon reflection what exactly physicalism is.

All are agreed that it is the view that fundamental reality is entirely physical: that the physical facts constitute the fundamental nature of the world. But there are two ways in which this initial definition cries out for clarification. Firstly what is it for a fact to be physical? Secondly, what is it for certain facts to entirely constitute fundamental reality? I will try in this chapter to answer both of these questions in turn.

Addressing the first question I take this to be a largely terminological issue; a matter of working out which use of the word ‘physical’ is best suited for allowing us to have the philosophical discussions we want to have. I won’t claim to have given the single best possible definition of physicality; and few if any of the significant conclusions of later chapters are dependent on my favoured definition. But this preliminary issue must be cleared up before we can get on to the substantive issues.

Addressing the second question is more substantive, as it takes us to basic issues concerning what metaphysical enquiry is all about. Metaphysicians are not just interested in a big list of all the things that exist: tables (tick), unicorns (X), God (TBC). For one thing putting together such a list would take a very long time. But more importantly, such a list wouldn’t tell us how everything ‘hangs together’; it wouldn’t reveal to us the unity lying behind the plurality of things which exist.

Metaphysicians want to dig deeper, down to the fundamental building blocks of reality. We are primarily interested not in what exists, but in what exists fundamentally; not in reality but in fundamental reality. Fundamental reality is the deep nature of the world; the metaphysical foundations upon which all being depends. In the second half I will articulate a not especially original view as to what this project amounts to.
Part I – The Nature of Physicality

A priori and a posteriori definitions of the physical

Contemporary debates between physicalists and their opponents have historical connections with debates in the 17th and 18th centuries between materialists and their opponents. Thomas Hobbes defended the view that fundamental reality was entirely material; in this he was opposed by George Berkeley who thought that fundamental reality was entirely immaterial, and by Rene Descartes who thought that fundamental reality was partly material and partly immaterial.

How these philosophers defined matter was shaped by, and itself shaped, the science of the time. Descartes thought of matter as extended stuff: stuff stretched out in three spatial dimensions. John Locke added solidity to Descartes’ characterisation, to differentiate corporeal things from souls which Locke thought might also have a spatial location. This roughly constituted the received understanding of material substance in the early days of the scientific revolution.

The trouble with these definitions is that twentieth century physics arguably shows that at least some physical stuff lacks these characteristics. Electrons are thought to be ‘point particles’, filling no space at all. So understood electrons lack extension, and hence by Descartes’ definition lack physicality (at least if we take ‘physical’ and ‘material’ to be synonymous, although I will later distinguish them). Modern physics talks about fields and particles, entities not naturally described as ‘solid’; and macroscopic objects are mostly empty space, which makes their classification as ‘solid’ in any pre-theoretical sense dubious (of course some objects are ‘solid’ by our modern chemical definition).

If physical stuff isn’t characterised by the properties of extension and solidity, which properties do define it? A natural thought is to look to physics to tell us what physical stuff is. It is twentieth century physics that has rendered the seventeenth century definition of matter obsolete; perhaps we can use modern physics to plug the hole it has itself created.

There are two ways of doing this. One way would be to abstract from our current physical picture of the world some general characteristics, and then stipulate necessary and
sufficient conditions for a property to count as ‘physical.’ To do this is to give an *a priori* definition of the physical. The worry about such an approach is that, as has proved to be the case with the 17th century definition of matter, as the science moves on the definition will become outdated. The alternative is simply to define the physical as whatever physicists tell us it is; to do this is to give an *a posteriori* definition of the physical.¹

Ultimately I will defend an *a priori* definition of the physical, but before I get to this I will explore the prospects for an *a posteriori* definition.

**Hempel’s dilemma and physics-based responses to it**

This is a much discussed problem with *a posteriori* definitions of the physical, which has become known in the literature as ‘Hempel’s dilemma’, due to its origins in the writings of Carl Hempel.² The dilemma arises when we ask ourselves whether in our definition of the physical as the subject matter of ‘physics’, we mean current physics with all its flaws, or perfect completed physics of the far off future. Both options have problems.

The problem with defining the physical in terms of current physics is that current physics is almost certainly false in at least some respects. Our best theory of the very big, i.e. general relativity, is inconsistent with our best theory of the very small, i.e. quantum mechanics. And the history of past physical theories being superseded by later physical theories gives us reason to suppose that current physics will one day be superseded by some more accurate theory of the universe. If physicalism is the view that all properties and states of affairs are grounded in the kinds of properties and states of affairs current physics talks about, then physicalism is almost certainly false.

There are a number of related difficulties involved in defining the physical facts in terms of future physics. Firstly there is a worry about *vagueness* (in the non-technical sense of that term). Who knows what weird and whacky entities future physicists will postulate? Without some way of narrowing down what future physics might look like it’s not clear what metaphysical views a commitment to ‘physicalism’ is supposed to rule out, and hence unclear what view one is signing up to when one claims to be a ‘physicalist.’ Some philosophers have expressed a similar worry about *vacuity*. If we just define physics as the

---

¹ This terminology is from Ney 2008a.
² Hempel 1966.
‘complete final theory of everything’, then it looks like a commitment to physicalism rules out nothing, since a physics which failed to include everything would by definition not be ideal. So defined physicalism is vacuously true.  

A final worry with future physics-based definitions concerns its potentially counterintuitive implications. Suppose future physicists postulate fundamental mentality, psychic powers, or even souls. It would then turn out that the truth of physicalism was consistent with the existence of fundamental mentality, which many philosophers find deeply counterintuitive. Physicalism is supposed to be the contemporary analogue of seventeenth century materialism, the arch opponent of dualism and idealism. If physicalism turns out to be consistent with such views then it seems to lose its point.

There are three options in the light of Hempel’s dilemma. We can adopt currentism: embrace the first horn of Hempel’s dilemma and try to define the physical in terms of current physics. We can adopt futurism: embrace the second horn of the dilemma and try to define the physical in terms of future physics, somehow dealing with the worries about vagueness, vacuity and counterintuitive implications. Or we can avoid Hempel’s dilemma altogether by returning to an a priori rather than an a posteriori definition of the physical.

As far as I know, the only philosopher who has gone for a pure form of currentism is Andrew Melnyk. Melnyk accepts that such a definition entails that physicalism is false. He nonetheless defends physicalism on the grounds that philosophers should aim at the view which is most probable. Even if physicalism is false, it may still be more probable than its competitors, such as dualism and idealism. Melnyk’s position is ingenious and more difficult to refute than you might think. However, given that one central aim of this book is to argue against physicalism, I will try to avoid beginning with a definition which rules out its truth from the off.

Janice Dowell is perhaps the best known defender of futurism; for Dowell the physical facts are whatever completed physics tells us they are. She answers the worries about vagueness and vacuity by putting a number of constraints on what counts as physics proper. This is

---

3 For a good survey of philosophical discussion surrounding Hempel’s dilemma see Ney 2008a.  
4 Melnyk 2003.  
5 Dowell 2006.  
6 Given these a priori constraints on what physics is, Dowell’s definition is not wholly a posteriori.
done in two ways. Firstly, physics is defined as our complete and ideal scientific theory of entities existing at relatively low levels of complexity. Completed physics then is not just ‘the complete theory of everything’, as was suggested in the concern about vacuity, it is the complete theory of things with relatively few parts. Electrons are currently thought to have no parts, and so would clearly be amongst the subject matter of physics under this definition. Protons are thought to be made up of two up quarks and a down quark, which gives them a bit more complexity, but this number of components is minuscule relative to the enormous number of parts which make up even a single cell (never mind a heart, a brain, or a whole organism).

Already we are starting to get content to the view. This stipulation ensures that the truth of physicalism rules out emergentism. Emergentism is the view that fundamental entities emerge at higher levels of complexity. 19th century British emergentists, for example, were committed to properties or states of affairs at the chemical, biological, and mental levels which were wholly irreducible to, and could not have been predicted from knowledge of, more basic properties and states of affairs.7 This commitment to properties/states of affairs over and above the properties/states of affairs at low levels of complexity renders emergentism inconsistent with physicalism, given Dowell’s definition. For the physicalist, fundamental reality is flat rather than layered.

Dowell puts further flesh on the bone by putting constraints on what kind of theories count as ‘scientific’. A scientific theory for Dowell must have the following four characteristics:

- **Testability** – The inclusion of a set of explanatory hypotheses from which empirically testable implications can be derived.
- **Variety** – Confirmation by the obtaining of a number and variety of the test-implications of its explanatory hypotheses.
- **Unity** – The provision of a unified explanation of a variety of empirical generalisations. The theory as a whole provides a unified explanation of the empirical generalisations that are amongst its testable implications.

---

7 For examples of British emergentism, see Mill 1843, Broad 1925, Alexander 1920. For a good discussion of British emergentism see McLaughlin 1992.
• Holistic – The theory receive additional empirical support by its fit with what is antecedently known and independently observable.

Having offered these constraints, Dowell then defines a ‘physical property’ as one which is ‘well-integrated into the most complete and unified explanation possible for the relatively most basic occupants of space-time.’¹⁸ This gives further response to the worries about vacuity and vagueness by ruling out certain possibilities. For an entity to be well-integrated into a scientific theory its behaviour must be regular and hence predicatable. Suppose that there are fundamental spirits which behave on whim, and hence are highly unpredictable. Such creatures could not be integrated into our best scientific theory, and hence their nature would not count as physical according to Dowell’s definition. The existence of such creatures is inconsistent with physicalism, given Dowell’s definition of the view.

What about the counterintuitive implications worry? Here Dowell bites the bullet. If physicists of the future end up attributing fundamental mental properties to particles, then the physical facts will involve fundamental mentality: physicalism will entail a form of panpsychism or idealism. How then do we make sense of physicalist’s traditional opposition to such views? Dowell thinks that we can avoid this concern by being careful not to conflate two distinct theses: its being highly improbable that physicalism will end up being consistent with panpsychism/idealism, and its being a priori incoherent that physicalism will end up being consistent with panpsychism/idealism. The physicalist can still remain opposed to panpsychism on the grounds that it is vanishingly improbable, according to Dowell, that future physicists will attribute mental properties to particles, and hence the truth of physicalism is highly likely to entail the falsity of panpsychism. It is overkill to insist that the truth of physicalism renders panpsychism not only highly improbable but a priori false.

Dowell’s contentment for the definition of physicalism to be consistent with panpsychism has not proved popular. Jessica Wilson offers a physics-based definition of physicalism very similar to Dowell’s, although incorporating reference to both present and future physics, but adds the following constraint:

**NFM Constraint** – The physical facts do not involve fundamental mentality.⁹

Given this stipulation, if physicists of the future do end up postulating fundamental mentality, physicalism will thereby be refuted. The NFM constraint restores physicalism to its position as traditional opponent of idealism, panpsychism and substance dualism.

Wilson ends up with the following definition of physicalism:

*The physics-based NFM account* – An entity existing at a world \( w \) is physical if and only if:

(i) It is treated, approximately accurately, by current or future (in the limit of inquiry, ideal) versions of fundamental physics at \( w \), and

(ii) It is not fundamentally mental (that is, does not individually either possess or bestow mentality)

Thus, Wilson’s definition is an interesting hybrid of a priori and a posteriori definition: it defers to physics, whilst retaining the a priori NFM Constraint. Note also that it’s deferral to physics combines both currentism and futurism.

In what follows I will examine a couple of further constraints on the nature of physics not considered by either Dowell or Wilson, the second of which I will suggest should serve as a constraint on the definition of physicalism. I will then go on to give a general objection to a posteriori definitions of physicalism, before finally defending an a priori definition.

**Pure physicalism**

One possible constraint not remarked upon by either Dowell or Wilson is the fact that, as discussed in chapter 1, from Galileo onward mathematics has been the language of physics. The maths has changed a great deal, but from Newton to Einstein to the present day there is a deep commonality in the fact that all physical theories are framed in the language of mathematics. In fact physics is not entirely mathematical; it also involves *nomic terms*, by which I mean terms expressing the concepts of *causation, natural necessity, or laws of nature*. Thus, we can say that physics has an entirely *mathematico-nomic* vocabulary.

---

This is potentially a serious constraint on physics, answering at least some of the worries about vagueness and counterintuitive implications. No matter how weird and whacky fundamental scientific theories of the future are, arguably no future scientific theory will count as a ‘physical’ theory if it does not restrict itself to a mathematical and nomic vocabulary. We worried above that future physicists might postulate fundamental mentality, or psychic powers. But in so far as these things have a nature that could not be captured in a mathematical or nomic terms, it will not be possible to capture them in the language of physics. Physics as it has existed since the scientific revolution has limited itself to describing the causal structure of reality, i.e. that which can be captured in a mathematico-nomic vocabulary, and we may reasonably take this to be a constitutive constraint on the subject.

What do I mean by talk of what can be ‘captured’ in a mathematical or nomic vocabulary? I do not simply mean what can be modelled or described with mathematical language. In the discussion of transparent and opaque concepts in chapter 1, I expressed the view that entities, or at least some of them, have natures (or ‘essences’, I use these words interchangeably), in the sense that it is a factual matter what the reality of the entity consists in. The best way to get a grip on this notion, and to make a case for it, is with reference to examples:

- The nature/essence of sphericity: For it to be the case that there is something spherical is for it to be the case that there is something with all points on its surface equidistant from its centre.
- The nature/essence of party-hood: For it to be the case that there is a party is for it to be the case that there are people revelling.
- The nature/essence of water: For it to be the case that there is water is for it be the case that there is something composed of H2O molecules.

We can call the above descriptions ‘metaphysical analyses’ of the entities in question; descriptions which capture what it is for the entity to be part of reality. When I say that an entity can be ‘captured’ in a mathematical or nomic vocabulary I mean that a metaphysical analysis of it can be given in that vocabulary; the property of sphericity would be an example of such an entity.
Restricting the physical to that which can be captured in a mathematical and nomic vocabulary is arguably too much of a constraint, as it entails that physicalism is a form of causal structuralism, the view that all there is to fundamental reality is causal structure. If physics only describes causal structure, and physics gives us a complete description of fundamental reality, it follows that all there is to fundamental reality is causal structure. We shall discuss causal structuralism in detail in chapter 6, but for now suffice to say that this is too austere a metaphysical picture for many philosophers, including many who self-identify as ‘physicalists’. I don’t think it’s helpful to class all those who find a world of pure causal structure unintelligible as enemies of physicalism. This is especially true if our main interest in defining physicalism is debating the mind-body problem, for many hardened reductionists about consciousness may be counted as ‘anti-physicalists’ simply because they think there’s more to reality than pure causal structure.

Nonetheless I do think it’s helpful to have a word for the especially pure form of physicalism according to which fundamental reality can be completely captured in the language of physics. Physicalists are philosophers who look to physics for their metaphysics. In the ideal form of this approach, physics has the linguistic resources to completely describe fundamental reality. Let us call physical facts which can be captured in the mathematico-nomic vocabulary of physics ‘pure physical facts’, and physicalism in conjunction with the view that fundamental reality wholly consists of such facts ‘pure physicalism’. It will greatly simplify my arguments against physicalism in the first half of the book to take pure physicalism as my target in the first instance, before going on to show in chapter 6 how they apply to physicalism in general.

**Naturalism and value-laden causal explanations**

Galileo made our fundamental science mathematical. For the reasons I gave above, I don’t think it is helpful to demand that the physicalist concur with Galileo in holding that physical reality can be exhaustively described in the language of mathematics. However, there was another crucial way in which Galileo changed the way we do science, and with regards to this second revolutionary change, I do think it reasonable to expect philosophers who call themselves ‘physicalists’ to be on board with it. I am thinking of the Galilean rejection of the teleological explanations of Aristotelian science.
Amongst other kinds of explanation, Aristotle favoured teleological explanations of substances in the natural world. The fact that a plant blossoms could be explained, for Aristotle, in terms of the *telos* of the plant: the set of goals which the plant, by its very nature, aims at. Similarly, the downward movement of the rock could be explained in terms of the telos of the earth the rock is made out of: earth, by its very essence, aims at getting to the centre of the universe (which Aristotle believed to be in the centre of the Earth).

A telos is an evaluative notion: what is good for an X is determined by X’s telos. The blossoming of the plant, the fulfilling of its telos, constitutes the plant’s good. The telos of man is *Eudemonia*, which is what constitutes the flourishing human life. Thus, teleological explanations, at least in the Aristotelian system, are value-laden explanations of causal happenings in the natural world.

Teleological explanations are highly controversial, but there is another kind of value-laden explanation which is common place: intentional explanations of human agents. All the time we explain the behaviour of people in terms of their responsiveness to reasons. When we say that Jimmy refused the job because it would have made him miserable, we are not attributing causal efficacy to a future counterfactual state of affairs: the non-obtaining state of affairs of Jimmy working in a miserable job. Rather, we explain what Jimmy did in terms of his recognition of and responsiveness to a certain reason; Jimmy recognises that the fact that the job would be miserable counts in favour of his turning it down, and he responds behaviourally in accordance with this reason.

The notion of a reason is constitutively connected to the notion of an ought: what ought to be done is determined by the balance of reasons. Some intentional explanations explain the agent’s action in terms of what she *took* to be a reason, but was not in fact a reason at all. Consider the following examples from Dancy:

His reason for doing it was that it would increase his pension, but in fact he was quite wrong about that.
The ground on which he acted was that she had lied to him, though actually she had done nothing of the sort.\textsuperscript{10}

In these cases, we get a motivating reason – the consideration which the agent acted for – without a normative reason – a consideration which counted in favour of action. The consideration for which the agent acted did not in fact obtain, and so did not count in favour of the action (as Dancy points out, motivating reason explanations are not factive). However, assuming a minimal realism about the normative, there are at least some cases in which motivating and normative reason go together; in such cases the agent recognises and responds to a good reason for action.

It is, I suggest, counter to the spirit of naturalism to take such value-laden explanations to be fundamental. As with the turn to mathematical theories, the rejection of teleology is a defining moment in the scientific revolution. Whilst the view that material reality is entirely mathematical is a radical view even amongst naturalists and physicalists, the Galilean rejection of value-laden explanations of the material world has been a widespread and long standing commitment of naturalistic philosophers. This is particularly clear with respect to the treatment of intentional explanations of action. Many naturalists deny that intentional explanations may be dispensed with, or analysed into, or superseded by, mechanistic explanations. Nonetheless, it is assumed that the capacity of an organism to behaviourally respond to reasons is in some sense constituted by more fundamental goings on, such that those more fundamental goings on can be understood mechanistically.

To make the point vivid, consider the hypothesis that there are angels with a primitive capacity to respond to reasons. The angel Gabriel sees that Joan is sick, recognises that that this gives him a reason to intervene, and miraculously heals her. Gabriel’s capacity to respond to reasons in this manner is not in any sense grounded in some more fundamental mechanism; he is an incorporeal being with a primitive capacity to act in response to reasons. Accepting the existence of such a being is, I suggest, counter to the spirit of naturalism.

\textsuperscript{10} Dancy 2000: 132.
The word ‘naturalism’, like the word ‘physicalism’ is slippery. Just as Dowell and Melnyk offer a posteriori definitions of physicalism, so we might offer a posteriori definition of naturalism: the natural truths are whatever the natural sciences tell us they are. We will then have to deal with worries about vagueness, vacuity and counterintuitive implications similar to those discussed above in the context of defining physicalism. Furthermore, below I will raise a general difficulty concerning a posteriori definitions of metaphysical positions. Better then to try to come up with an a priori definition of ‘naturalism’ which captures the general view of most philosophers who self-identify as naturalists, so that we can get on with debating whether or not such philosophers have a view which matches reality.

However we do end up defining naturalism, it ought to involve a commitment to the view that the fundamental causal workings of reality do not involve value. This does not entail that naturalists must deny the reality of value and humans’ responsiveness to it; but it does impose on naturalists an obligation to explain the capacity of humans to respond to value in more fundamental terms. This fits with the aspirations of those who call themselves ‘naturalists’ in both metaethics and the philosophy of mind. And it chimes with the rejection of value-involving causal explanations in the physical sciences from the scientific revolution onwards. This is a metaphysical view suited to those who look to the sciences for their metaphysics, although the definition itself does not defer to the sciences (and thus is an a priori rather than an a posteriori definition).

In line with this, I think it appropriate to consider physicalism as a specific form of naturalism. In general the naturalist looks to the natural sciences to find out the fundamental nature of the world. The physicalist looks specifically to our most general and basic science. Emergentism could be thought of as a form of naturalism, but one which postulates fundamental laws at higher levels, such as the chemical, biological or mental level. The physicalist need not deny the reality of higher-level laws or properties, but she must deny that they are part of fundamental reality.

Finally we get to my first concern with Dowell and Wilson, which is that neither account rules out value having a role to play in the fundamental causal workings of reality. Dowell’s account of a scientific theory places emphasis on regular behaviour which is predicatabale. However, there is no a priori reason why value-involving causal explanations could not yield
highly predictable behaviour. Suppose that particles have a fundamental telos for constituting living things, a view for which Thomas Nagel has recently expressed sympathy.¹¹ It could be that the behaviour of a particle is highly predictable on the basis of this telos in conjunction with the laws of physics. Nonetheless, contemplating the existence of such laws seems contrary to spirit of physics-based metaphysical enquiry.

Dowell may respond here just as she did to the concern that future physicists will postulate fundamental mentality: physicalists need not reject value-laden causal explanations as inconsistent with their view; they can simply oppose them as highly improbable. However, the rejection of value-laden explanations seems just as plausible a constitutive condition on physics as the four constraints Dowell herself supports. From Galileo onwards, physical theories have been framed in the value-neutral language of mathematics and non-normative causation.

It is hard to see how Wilson could evade this concern, as fundamental value seems no less at odds with the spirit of physicalism than fundamental mentality. She claims in defence of her definition that all the entities we intuitively take to be inconsistent with physicalism – free will, intentionality, aesthetic value – are dependent on mentality, and hence covered by the NFM Constraint. However, fundamental natural teleology need not be dependent on fundamental mentality, and hence is not ruled out by the NFM Constraint.¹² The inconsistency of physicalism with fundamental value-laden laws must be separately stipulated.

This is not a fundamental criticism of either view; a small adjustment could be made to ensure that ‘the physical’, by definition, does not involve value-involving causal goings on. Dowell could incorporate a rejection of value-laden explanations in the definition of science proper as follows:

- **Testability** – The inclusion of a non-value involving set of explanatory hypotheses from which empirically testable implications can be derived.
- **Variety** – Confirmation by the obtaining of a number and variety of the test-implications of its explanatory hypotheses.

¹¹ Nagel 2012.
¹² For an articulation of the possibility of teleological causation see Hawthorne & Nolan 2006.
- **Unity** – The provision of a unified explanation of a variety of empirical generalisations. The theory as a whole provides a unified explanation of the empirical generalisations that are amongst its testable implications.

- **Holistic** – The theory receive additional empirical support by its fit with what is antecedently known and independently observable.

However, I think there are deeper problems with any physics-based definition of the physical, and it is to this deeper problem I now turn.

**Against a posteriori definitions of the physical**

On intriguing possibility for avoiding the difficulties associated with Hemple’s dilemma, whilst nonetheless tying physicalism very closely to physics, is given by the ‘attitudinal’ conception of physicalism defended by Alyssa Ney. On this view ‘physicalism’ does not name a metaphysical view of any kind, but rather an *attitude* to metaphysical enquiry: a physicalist is someone who, as it were, takes an oath to formulate her ontology solely according to the current posits of physics. When I made my non-negotiable commitment to the Consciousness Constraint in the last chapter, I was in effect taking a different vow: to ensure that my metaphysical view has a place for phenomenal consciousness. Even if I came to accept that the ontology postulated by physics happens to suffice for the satisfaction of phenomenal concepts, this fact would not make me an attitudinal physicalist; for I would be content with that ontology not simply because it was postulated by physics, but because it sufficed for the reality of phenomenal consciousness. To commit to the Consciousness Constraint is reject attitudinal physicalism.

Therefore, I will continue to think of physicalism as a metaphysical view, as a view about how fundamental reality is independently of how we happen to take it to be. And understood as such, it is prima facie strange to define physicalism with reference to an epistemological activity of humans, no matter how important we think that activity is to finding out about the world. We are not omniscient beings, we learn about the world through fallible methods, and it is clearly possible that fundamental reality might be different, perhaps in subtle respects, from the picture of reality we get from the best use of

---

13 Ney 2008b. See also van Fraassen 2002.
those methods. Of particular relevance to our current enquiry, the world might turn out
different from the picture of reality yielded by ideal physical theory.

My concern is not that physicalism might be false. Few physicalists think that we can be \textit{a priori} certain of the truth of physicalism, so it is not a problem if the definition of the view is
compatible with its falsity (it would be a problem if it weren’t). The worry is that ideal
physics might get fundamental reality a bit wrong \textit{even if physicalism is true}.

Consider the following example to help focus the issue. Mass is an important property in
physics. Suppose there were an epiphenomenal property, i.e. a property with no causal
powers, which was always co-located with mass. Call that property ‘shadow mass.’ On this
supposition, all objects constituted of mass are equally constituted of shadow mass.
Physicists are interested in mass because of its dynamical effects: gravitational attraction
and resistance to acceleration. But given that shadow mass is epiphenomenal, it’s not going
to show up even in our best physics. An ideal physics could perfectly predict the behaviour
of all entities in space and time without ever having need to postulate shadow mass. And if
it never has need to postulate shadow mass it never will. Nonetheless it could be there.
There is no \textit{a priori} reason to suppose that reality can’t outstrip our best theories.

I take it that there could be shadow mass, and we are imagining for the sake of the example
that there is. On the physics-based definitions of Dowell and Wilson, shadow mass is not a
physical property, as it doesn’t show up in current of complete physics. Physicalism is
therefore false, as there is a non-physical fundamental property. And yet intuitively the
existence of shadow mass is not enough to refute physicalism. Philosophers count
themselves as physicalists in virtue of their passionate opposition to emergentism, idealism,
panpsychism, substance dualism, theism, belief in demons and poltergeists. Physicalists do
not get irked by a property which is no more mental or supernatural or otherwise
extraordinary than mass. Intuitively, in the shadow mass hypothesis, reality is wholly
physical; it’s just that fallible human physics gets physical reality a tiny bit wrong.

Perhaps some physicalists might feel that the existence of epiphenomenal properties is
counter to the spirit of physicalism. But we could easily cook up a different scenario in which
reality differs a bit from how ideal physics takes it to be, one not involving epiphenomena.
Physicists don’t just read off theories form the empirical data; they are guided in theory
choice by principles of theoretical virtue. Physicists aim at the simplest, most economical account of the empirical data, and I assume that it is rational for them to do so. But this is not an infallible method: it could turn out that reality itself is not as simple as the simplest interpretation of the empirical data takes it to be, in which case reality will differ from how ideal physics takes it to be. For example, it could turn that the simplest interpretation of the empirical data postulates four fundamental forces – which entails that ideal physics postulates four fundamental forces – whilst reality itself involves five fundamental forces.

Wilson predicts something like my concern and offers the following response:

Some [e.g. Dowell] might respond...by accepting that physicalism would be falsified in such a scenario [i.e. one in which fundamental reality deviates from what ideal physics tells us about it]. I prefer rather to put such sceptical possibilities aside, as failing to take the appeal to physics in the proper metaphysical spirit. This appeal is to be understood sufficiently generally that it provides a basis for a contentful, testable appropriately flexible formulation of physicalism...It is not also required that it provide such a basis in the face of every sceptical scenario, whether than involves brains in vats, insuperable cognitive limitations, or entities that are in-principle inaccessible.14

Contra Wilson, it doesn’t seem to me appropriate to describe the scenario in which complete physics gets things a bit wrong as ‘skeptical’. A working scientist is entitled to assume that she is not a brain in a vat or the star of the Truman Show, if only because science can’t get going unless we rule out such possibilities. But she is not entitled to assume that her science will inevitably get the world exactly right. The possibility that physics may, even in its ideal form, get things a bit wrong is one proper epistemic humility, proper appreciation of human fallibility, requires us to give some credence to.

Given this, Wilson’s defence seems to be that her view is usable, but either a bit incomplete or a bit wrong. It’s a bit incomplete if it has nothing to say about the shadow mass scenario.

14 Wilson 2006: 74
It’s a bit wrong if it rules the truth of physicalism inconsistent with the shadow mass scenario. Why not do better if we can?

Wilson goes on to offer reasons why we may not be able to do better than an a posteriori definition of physicalism:

‘...if the characterization of the foundation is to go beyond the bare description of these as existing at relatively low orders of complexity...we have little choice but to appeal to physics.’

However, in the light of my response to the first concern above, we now have more resources with which to construct an a priori definition of the physical. We can think of naturalism as the view that fundamental reality does not involve value-laden causation, and we can think of physicalism as a special form of naturalism according to which fundamental reality consists entirely of entities at relatively low levels of complexity.

Definitions of physicality and materiality

What about the NFM Constraint? I think at this point we need to confront head on question of what we’re trying to do here. Physicalism isn’t some entity out there in the world which we are trying to discover the essence of. The word ‘physicalism’ is a technical term used by philosophers. In trying to define it we are either trying to track how philosophers happen to have used this term, which seems to me a project of limited interest, or we are trying to shape a definition which is useful for practitioners of philosophy.

What then are our interests in trying to define physicalism? They seem to me twofold. Firstly, we want to capture a certain kind of philosophical view which, inspired by the physical sciences, tries to explain the world in terms of non-value involving facts at relatively low levels of complexity. Secondly, we want to capture one side of the central debates in the mind-body problem: the side that opposes dualism, idealism and panpsychism; the side that thinks the knowledge argument and the conceivability argument are unsound; the side that thinks we don’t need to postulate special entities in order to explain mentality.

Without the NFM Constraint we can only fulfil the first of these needs. Many panpsychists hold that fundamental reality consists only of non-value involving states of affairs at relatively low levels of complexity, and yet are on the side of the debate opposite from the one we like to designate with the word ‘physicalism’. In fact, as will become clear in chapter 6 when we encounter Russellian monism, in order to really capture the two sides of the central debate in the mind-body problem, we need to go beyond Wilson’s NFM Constraint and stipulate that physicalism is inconsistent with fundamental mentality and proto-mentality (I will offer a definition of proto-mentality in chapter 6).

Therefore, I choose to define physicalism as follows:

*Physicalism* – Fundamental reality is wholly constituted of facts which (i) concern spatio-temporal entities at relatively low levels of complexity, (ii) do not involve value-laden causation (iii) do not involve mentality or proto-mentality.\(^{16}\)

Accepting this definition of ‘physicalism’ requires us to give a corresponding definition of a ‘physical fact’. In fact, it is useful to have both a broad and a narrow definition:

*Narrowly physical fact* – F is a narrowly physical fact iff F is a fundamental fact which (i) involves spatio-temporal entities at relatively low levels of complexity, (ii) does not involve value-laden causation (iii) does not involve mentality or proto-mentality.

*Broadly physical facts* – F is a broadly fact iff F is wholly grounded in a narrowly physical fact(s).\(^{17}\)

We can further define narrowly physical individuals and properties as the individuals and properties involved in narrowly physical facts, and broadly physical individuals and properties as the individuals and properties involved in broadly physical facts.

If we accept my definition of physicalism, then we could not simply define the physical as the subject matter of physics, because the entities physics refers to may turn out to involve mentality or value-laden causal powers. However, it is useful to have a term for facts

---

\(^{16}\) One could adjust the definition to avoid quantifying over facts. However, I am not concerned in this book with the issue of whether properties or facts are fundamentally real, and so I will continue to quantify over them for the sake of ease of exposition.

\(^{17}\) The terminology of ‘broadly’ and ‘narrowly’ physical is introduced in Montero 2013.
concerning individuals and properties at relatively low levels of complexity, regardless of whether or not they involve mentality or value-laden causation. I will refer to such facts as ‘material’, with the one stipulation that the individuals and properties they concern are mind-independent (so that the existence of matter is by definition inconsistent with idealism). So ‘matter’ is the stuff physics tells us about, and matter is ‘physical’ so long as it doesn’t involve fundamental mentality/proto-mentality/value-laden causation.

As in the case of physicality, we can distinguish between a narrow and a broad sense of materiality:

*Narrowly material facts* – F is a narrowly material state of affairs iff F is a fundamental fact which concerns spatio-temporal entities at relatively low levels of complexity.

*Broadly material states of affairs* – F is a broadly material fact iff F is wholly grounded in a narrowly material fact(s).

(Important qualification: In chapter 9 we will consider *priority monism*, the view that all facts are grounded in facts about the cosmos. As I will explain then, I take priority monism to be anti-emergentist, and hence a form of materialism. Forms of priority monism which deny the fundamental reality of (proto)mentality or value-laden causation I take to be forms of physicalism. Hence, at that stage of the book I will extend the definition of materialism and physicalism to cover priority monism. Until that point, I will work with the above micro-level focused definitions of physicalism and materialism for the sake of simplicity).

**Part II – The Nature of Fundamentality**

Now we have a grip on what physicality is, we turn to the question of what it would be for fundamental reality to be entirely physical. This calls for an account of fundamentality, of what it is for certain facts to entirely constitute fundamental reality. In the last chapter I introduced the notion of *grounding*. In what follows I will say a little more about how I understand grounding, outline a conception of physicalism defined in terms of grounding, and then defend it against alternative accounts of fundamentality.
Constitutive grounding and the Free Lunch Constraint

I follow the general line in the literature of taking grounding to be a non-causal explanatory relationship that obtains between facts or other entities.\(^{18}\) As with the notion of essence, the best way to clarify and make a case for the notion of grounding is with reference to examples. Suppose Rod, Jane and Freddy are dancing, drinking and generally having fun one evening at Jane’s. It follows from this supposition that there is a party at Jane’s, and moreover that there is a party at Jane’s because Rod, Jane and Freddy are dancing, drinking, etc. at Jane’s. But the word ‘because’ here does not express a causal relationship; the revelling does not causally bring into being the party. Consider a further example. Suppose the rose is scarlet. It follows that the rose is red, and moreover that the rose is red because it is scarlet. But the scarlet colour of the rose does not cause it to be red; the rose does not secrete redness as the liver secretes bile. It seems that in both cases we have a non-causal explanatory relationship; and this we call ‘grounding’.\(^{19}\)

As I said in the previous chapter, I take grounding to be non-causal explanatory relationship which obtains between facts or other entities.\(^{20}\) In that chapter I did little more than propose that we can get a grip on the notion with reference to examples:

- The fact that there is a party is grounded in the fact that Rod, Jane and Freddy and revelling.
- The fact that the rose is red is grounded in the fact that the rose is scarlet.

---

\(^{18}\) There is much debate about what the relata of grounding are. As regards grounding by analysis, I am attracted to the view (Fine 2001, 2012) that the logical form of grounding claims are \(<X \text{ because } Y>\), in which X and Y are sentences. On this view, grounding is not strictly speaking a relation. However, for ease of exposition I will generally think of grounding by analysis as a relationship between facts. In terms of grounding by subsumption (introduced in chapter 9), I am inclined to think of it as a relation which any kind of worldly entity – individuals, properties, states of affairs – can stand to any other kind of worldly entity. For more discussion on this, and other general issues regarding the nature of grounding, see Trogdon 2013a.

\(^{19}\) Some key papers on the recent revival of grounding are Fine 2001, Schaffer 2009a and Rosen 2010. Proponents of grounding trace the idea back to an older tradition, often citing Aristotle as an influence. In fact, I work with the more specific notion of constitutive grounding, as explained in chapter 2, but I will overlook that detail for the moment.

\(^{20}\) There is much debate about what the relata of grounding are. As regards grounding by analysis, I am attracted to the view (Fine 2001, 2012) that the logical form of grounding claims are \(<X \text{ because } Y>\), in which X and Y are sentences. On this view, grounding is not strictly speaking a relation. However, for ease of exposition I will generally think of grounding by analysis as a relationship between facts. In terms of grounding by subsumption (introduced in chapter 9), I am inclined to think of it as a relation which any kind of worldly entity – individuals, properties, states of affairs – can stand to any other kind of worldly entity. For more discussion on this, and other general issues regarding the nature of grounding, see Trogdon 2013a.
• The fact that there is a table in location L is grounded in facts concerning the arrangement of atoms in L.

Is there anything more we can say about the grounding relationship? A striking feature of many grounding relationships is that there is an intuitive sense in which the grounded fact/entity is *nothing over and above* its ground. Focus on the party example. It’s not as though there are the people dancing, drinking, etc., and then there’s this extra thing – the party – which floats above their heads. There’s a very intuitive sense in which the fact that there is a party is *nothing more* than the fact that there are people revelling; a world in which there are people revelling is *already thereby* a world in which there is a party. Or to turn to the colour example, it is not as though the rose has two distinct colours: its scarlet colour and its red colour. There’s a very intuitive sense in which the rose’s being red is nothing more than its being scarlet; a world in which the rose is scarlet is *already thereby* a world in which the rose is red. I shall call such grounding relationships – in which the grounded fact is nothing over and its ground – ‘constitutive grounding’ relationships.

It is not obvious that all grounding relationships are constitutive grounding relationships. G. E. Moore believed that goodness was a fundamental property in its own right, although one which supervened on the non-normative facts. Perhaps we ought to construe this a non-constitutive grounding relation, in which the facts about goodness are grounded in, but ontologically additional to, the non-normative facts. Nonetheless, physicalism is generally understood to be the view that all facts are nothing over and above the physical facts, and hence I will define it in terms of constitutive grounding. (More importantly, as we shall discover in chapter 6, it is only via the postulation of constitutive grounding relationships that physicalism – and indeed Russelian monism which we will consider in that chapter – has hope of offering a solution to the causal exclusion problem). In general, I will use the word ‘grounding’ to mean constitutive grounding.

Is there anything more we can say about the ‘nothing over and above’ relationship? I’m inclined to think that we are *obliged* to say something more; and that’s because the ‘nothing over and above’ relation has a prima facie paradoxical nature. On the one hand it seems almost tautological that if x is not *identical* with y, then x is *something over and above* y; from which it would follow that x can be nothing over and above y only if x is identical with
y. On the other hand, when we reflect on cases it is evident that the nothing over and above relation holds between non-identical facts. Recall the party example discussed above: the fact (F1) that there is a party is nothing over and above the fact (F2) that Rod, Jane and Freddy are revelling. We can see that (F1) and (F2) are distinct facts, as they can obtain independently of each other: if Rod and Freddy leave the party and are replaced by Peter and Solomon, then (F1) will obtain in the absence of (F2).

Thus we have a prima facie paradox: it seems that the nothing over and above relation must be the identity relation, and yet it’s clearly not the identity relation. And hence, if we want to indulge in ‘nothing over and above’ talk, we are obliged to clarify the nature of the constitutive grounding relation, in such a way that we have an understanding of how it could possibly be that x is nothing over and above y, despite the fact that x is not identical with y. I shall call this the ‘Free Lunch Constraint’, after David Armstrong’s famous term ‘ontological free lunch’ for an entity which is nothing over and above already postulated facts. In the next section I will outline a conception of constitutive grounding which satisfies this constraint: grounding by analysis.

**Grounding by Analysis**

In this section I will outline grounding by analysis and suggest that it satisfies the Free Lunch Constraint, and thus counts as an adequate form of constitutive grounding. I do not think that grounding by analysis is the only form of constitutive grounding; in chapter 9 I will outline an alternative form of constitutive grounding which features large in the metaphysical view I ultimately want to defend. However, grounding by analysis is a widely applicable ‘all purpose’ form of constitutive grounding; and hence in general I will understand physicalism (and other views defined in terms of constitutive grounding, e.g. constitutive Russelian monism, which we will be exploring from chapter 6 onwards) in terms of it.

In the discussion of pure physicalism above, I expressed the view that some entities permit of metaphysical analysis, in the sense that we can describe what the reality of the entity consists in, or equivalently what is essentially required for the entity to be real. For example, for sphericity to be instantiated it is essentially required that there is something with all points on its surface equidistant from its centre.
Grounding by analysis depends on this idea. It can be defined as follows:

Fact X is grounded by analysis in fact Y iff:

- X is grounded in Y, and
- Y logically entails what is essentially required for the entities contained in X to be part of reality.

To take a concrete example, the fact (F1) that there is a party is grounded by analysis in the fact (F2) that Rod, Jane and Freddy are revelling because:

- F1 is grounded in F2, and
- The fact that Rod, Jenny and Freddy are revelling logically entails what is essentially required for a party to exist, i.e. that there are people revelling.

I submit that grounding by analysis satisfies that Free Lunch Constraint. In cases of Grounding by Analysis, the grounding fact provides all that is essentially required for the entities contained in the grounded fact to be part of reality, e.g. the fact that Rod, Jane and Freddy are revelling provides all that is required for there to be a party. This gives us a clear sense in which the grounded fact is nothing over and above the grounding fact, even in the absence of identity between the facts or their constituents.

Grounding by analysis is in the first instance a relationship between facts. But we can think of the grounding of individuals and token properties as derivative on the grounding of analysis of facts: individual or property token E is grounded in fact F1 just in case E is a constituent of fact F2 and F2 is grounded in F1.

A grounding account of physicalism

An obvious first attempt at defining physicalism in terms of grounding would be the following:

---

21 Strictly speaking I prefer the sentential connective view expressed in footnote 18, but for ease of exposition we can think of grounding by analysis of a relation between facts.

22 The grounding by analysis model is influenced by the Fine/Dasgupata ‘grounding via essence’ I outline in the next section. However, there are important differences, as I explain below. The account is also similar to that of Andrew Melnyk, discussed below.
Strong Physicalism – Physicalism is the view that all facts which are not narrowly physical facts are constitutively grounded in the narrowly physical facts.23

This gives us a clear and straightforward way of understanding the view that fundamental reality is wholly physical: the many and diverse facts which make up reality are all ultimately grounded in the physical facts. However, Shamik Dasgupta has recently raised problems with Strong Physicalism, and urged that it be modified. His concerns arise from his conception of how cases of grounding are to be explained; a conception which is very close to the grounding by analysis model outlined above. (This is no coincidence, as my formulation of it was influenced by Dasgupta’s work, although there are important differences, as I explain in footnote 27.) We shall now explore his concerns.

Suppose that Sarah is currently feeling pain. Strong Physicalism commits the physicalist, as we would expect from a definition of physicalism, to there being some physical fact which grounds the fact that Sarah feels pain. Let us suppose, to continue with the hackneyed and empirically dubious example favoured by philosophers, that that physical fact is the fact that Sarah’s c-fibres are firing. Thus we reach the following fact:

\[
\text{Pain-Grounding} \quad \text{– The fact that Sarah’s c-fibres are firing grounds the fact that Sarah feels pain.}
\]

Pain-Grounding is a grounding fact, that is to say, a fact about which facts ground which. In formulating his grounding conception of physicalism, Dasgupta of course accepts that the physicalist is obliged to hold that the facts about consciousness are grounded in the physical facts, and hence is obliged to accept grounding facts similar to Pain-Grounding. However, he denies Strong Physicalism because he does not think that the physicalist needs to hold that the grounding facts themselves, i.e. facts like Pain-Grounding, are wholly grounded in the physical facts.

If Strong Physicalism is true, then Pain-Grounding, like any other fact, is grounded in the physical facts. But Dasgupta argues that Pain-Grounding cannot be satisfactorily explained in terms of the physical alone; rather it must be explained at least in part in terms of the

23 Dasgupta does not use the terminology of ‘narrowly’ and ‘broadly’ physical; he uses the word ‘physical’ to mean what I call ‘narrowly physical’.
nature of pain. It is truths concerning the nature of pain, i.e. what it is for something to feel pain, which explain why it is that the firing of c-fibres grounds pain.\textsuperscript{24}

To make this plausible, return to our party example. Just as Pain-Grounding concerns the grounding of pain, so the following fact concerns the grounding of parties:

\textit{Party-Grounding}: The fact that Rod, Jane and Freddy are revelling grounds the fact that there is a party.

Why is it the case that the fact that Rod, Jane and Freddy are revelling grounds the fact that there is a party? Intuitively this is because of the nature of a party, because of what a party is: a party is the kind of thing that exists when there are people revelling. In this way the nature of a party ‘opens itself up’ to the possibility of being grounded in specific facts concerning revelling.

Dasgupta proposes we explain Party-Grounding in terms of the following two facts:

\textit{Party-Nature} – A party is essentially such that if there are people revelling then there is a party,

\textit{Revelling} – Rod, Jane and Freddy and revelling.\textsuperscript{25}

Note that the entities in the \textit{less fundamental fact} – the fact that there is a party – are doing crucial explanatory work in the explanation of the overall grounding fact. Dasgupta argues, partly through reflection on cases, that we do not get a satisfying explanation of grounding facts from the more fundamental fact alone. For example, it would not be satisfying to answer:

“Why is it the case that the fact that Rod, Jane and Freddy are revelling grounds that fact that there is a party?”

\textsuperscript{24} Dasgupta also suggests that we might also explain grounding facts in terms of conceptual truths or metaphysical laws, but does not outline these proposals in detail. The former alternative would seem to lead to difficulties similar to the difficulties I explore below concerning attempts to account for fundamentality in terms of representations of reality, rather than in terms of worldly entities. The latter model seems to me not very promising, as metaphysical laws are intuitively the kind of things we want to explain. Dasgupta’s discussion starts from a problem Sider (2012) raises with grounding theories of fundamentality, but it would be distracting to explore that here.

\textsuperscript{25} Dasgupta’s example is in terms of conferences rather than parties, but the substance is the same.
with:

“Because Rod, Jane and Freddy are revelling”

It is only by reference to the nature of parties, to what a party is, that we get a satisfactory explanation of Party-Grounding.Kit Fine has previously advocated a similar kind of ‘top-down’ direction in the explanation of grounding facts:

...what explains the ball’s being red or green in virtue of its being red is something about the nature of what it is for the ball to be red or green (and about the nature of disjunction in particular) and not something about the nature of what it is for the ball to be red. It is the fact to be grounded that ‘points’ to its grounds and not the grounds that point to what they may ground.

It is not that the less fundamental fact ‘points to’ the specific facts which ground it; essential truths concerning parties do not involve specific reference to Rod, Jane and Freddy. Rather the nature of constituents of the less fundamental fact F ‘point to’ some condition which is sufficient for its being the case that F.

Thus, Dasgupta proposes explaining Pain-Grounding in the following way:

\textit{Pain-Nature} – Pain is essentially such that for any x, if x’s c-fibres are firing then x feels pain.

\textit{C-Fibres} – Sarah’s c-fibres are firing.

Let us call this this view that grounding facts are explained in terms of the nature of constituents of the grounded fact ‘the grounding via essence’ model. There is a tension between the grounding via essence model and Strong Physicalism: Strong Physicalism

---

26 Karen Bennett (2011) and Louis deRosset (2013) try to ground the grounding facts in the fundamental facts. I find their arguments unconvincing, but it would be distracting to get into the discussions here.

27 Fine 2012.

28 Again, Dasgupta examples are slightly different to mine – involving consciousness rather than pain – but the substance is the same.

29 Note that I am interpreting the grounding via essence model such that it is the essence of constituents of the grounded fact that explains the overall grounding fact. Rosen (2010) considers a more permissive model according to which the overall grounding facts is explain either in terms of the grounded fact or in terms of the fact which does the grounding.
maintains that all facts are grounded in the fundamental physical facts, whilst according to
the grounding via essence model the grounding facts are partly grounded in facts about the
nature of higher-level entities and so are not wholly grounded in the fundamental physical
facts. Pain-Grounding, for example, is partly grounded in Pain-Nature. The grounding via
essence model entails that chains of grounding explanation don’t always move in a
downward direction. For this reason Dasgupta rejects Strong Physicalism, and adopts a
definition of physicalism according to which certain facts are ‘exempt’ from needing to be
grounded in the physical, even if physicalism is true.

I am not persuaded that the grounding via essence model applies to all cases of grounding.
It fits cases of grounding by analysis, in which the metaphysical analysis of the grounded
entities plays a key role in accounting for the grounding. In chapter 9 I will outline an
alternative form of constitutive grounding which works in a quite different way, and hence
does not fit the grounding via essence model.

I am also not persuaded that the grounding via essence model, in itself, satisfies the Free
Lunch Constraint, because of counterexamples suggested to me by Adam Pautz.30 We briefly
discussed earlier the G. E. Moore style view according to which goodness is a fundamental
property in its own right, but nonetheless the facts about goodness are grounded in the
non-normative facts. This would be a case of non-constitutive grounding: the facts about
goodness are grounded in, but are ontologically additional to, the non-normative facts. Now
the proponent of this view may very well hold that such grounding facts fit the grounding
via essence model. They could hold that goodness is essentially such that if an act maximises
wellbeing then it is good – even though goodness and wellbeing are wholly distinct
properties – which would explain the fact that facts about wellbeing ground facts about
goodness.

Or, to consider a case closer to our concerns, we might imagine an analogous view of the
relationship between the mental and the physical. An emergentist might hold that pain is a
fundamental feature of reality, but accept Dagupta’s model of the grounding of pain: it is in

30 In conversation, but Pautz discusses one such counterexample in Pautz MS: 34-35.
the nature of pain that if someone has their c-fibres firing then they feel pain. We would thus have a case of non-constitutive grounding of the mental in the physical.

The grounding by analysis model seems to me to avoid these concerns. It fits the grounding via essence model, but it is less permissive. If the fact that action A maximises wellbeing logically entails *what it is for goodness to be instantiated* – which can presumably be the case only if all it is for an action to be good is for it maximise wellbeing – it follows that the resulting instantiation of goodness is nothing over and above the fact that A maximises wellbeing. Similarly if the fact that Sarah’s c-fibres are firing logically entails what it is for pain to be instantiated – which can presumably be the only if all it is for someone to be in pain is for their c-fibres to be firing – it follows that the resulting instances of pain is nothing over and above the fact that Sarah’s c-fibres are firing.

Nonetheless, so long as there are cases of grounding which fit the grounding via essence model – and cases of grounding by analysis are a subset of such cases – we need a conception of physicalism which allows for them. Strong Physicalism, therefore, must be rejected.

Obviously there must be some limit on which facts physicalism ‘allows’ not to be grounded in the physical; it is inconsistent with physicalism, for example, to deny that the facts about consciousness are grounded in the physical. Dasgupta’s view is that physicalism does not require that *facts about natures* are grounded in the physical. This is because, according to Dasgupta, facts about natures are *autonomous*, or not apt to be grounded; that is to say they are not the kind of fact for which the question of grounding arises.

Dasgupta offers an analogy to help clarify and motivate the thesis that facts about natures are autonomous. The analogy is between facts which are not apt for grounding and facts which are not apt for causal explanation. The fact that 2+2=4 lacks a causal explanation, but not in the sense that the big bang may lack a causal explanation; the fact that 2+2=4 is not the kind of fact which requires or admits of causal explanation. By analogy there may be a category of fact which neither requires nor admits of grounding explanation, and essential truths are a plausible candidate. According to Dasgupta, the question ‘What explains the

---

31 The argument for the autonomy of facts about natures is continued in Dasgupta 2014.
fact that a party is the kind of thing that exists when there are people revelling?’ is ill-posed in something like the way ‘What caused 2 and 2 to equal 4?’ is ill-posed. Nobody who knows what a party is should be troubled by this question.

In the light of this, Dasgupta offers the following improvement on Strong Physicalism:

> Weak Physicalism – Physicalism is the thesis that all facts which are not narrowly physical facts and which are substantive, i.e. apt to be grounded, are constitutively grounded in facts which are either narrowly physical or autonomous.

He does not ultimately settle with this definition, as so defined physicalism is consistent with the existence of God and Platonistic numbers, so long as these things are wholly grounded in autonomous facts about their natures. In order to avoid this worry, he ends up with the following definition:

> Moderate Physicalism – (i) Weak Physicalism is true, and (ii) all autonomous facts help underwrite the kind of grounding explanations required by Weak Physicalism.

Dasgupta’s final definition seems to me too strong. The physicalist can surely accept that there are non-physical properties and individuals in other possible worlds, which would seem to require that there are autonomous facts concerning unactualised non-physical natures. And surely such autonomous facts have little to do with grounding explanations in a purely physical world.

We can avoid the worry by stipulating that the physicalist denies that there are any individuals, properties or events which exist or are instantiated wholly in virtue of autonomous facts, or perhaps more cautiously that the only individuals, properties or events which exist wholly in virtue of autonomous facts are narrowly physical individuals, properties or events.\(^{32}\) We can efficiently ensure this by making the following modification:

> Reasonably Moderate Physicalism – Physicalism is the thesis that all facts which are not narrowly physical facts and which are substantive, i.e. apt to be grounded, are

---

\(^{32}\) Elsewhere Dasgupta (2014) defends a version of the Principle of Sufficient Reason, which entails that there are entities which exist wholly in virtue of autonomous facts, and takes this to be consistent with physicalism.
wholly constitutively grounded either in narrowly physical facts alone, or in facts which are autonomous and in narrowly physical facts.

This modification removes the possibility of a God whose existence is grounded in facts about her nature, as the existence of such a being would not be grounded in the narrowly physical facts.

Thus, the definition of physicalism I will work with throughout this book is as follows:

**Physicalism** – All substantive facts which either concern entities at greater than atomic levels of complexity or involve mentality/proto-mentality or value-laden causation, are wholly constitutively grounded either in facts which concern entities at atomic or lower levels of complexity and which do not involve mentality/proto-mentality or value-laden causation, or in such facts in conjunction with autonomous facts.

I will not be concerned in this book with non-constitutive forms of grounding; hence from now on I will simply use the word ‘grounding’ with the understanding that it refers to constitutive grounding. And for most of the book I will understand constitutive grounding as grounding by analysis. If we want to adopt a form of constitutive grounding distinct from grounding by analysis, then we must find an alternative way of satisfying the Free Lunch Constraint. I will attempt to do this in chapter 9; until then I will equate constitutive grounding with grounding by analysis, referring to both as simply ‘grounding’.

For the rest of this chapter I will consider alternative conceptions of physicalism based on alternative accounts of fundamentality.

**Alternatives to grounding accounts of fundamentality**

**Accounts of fundamentality in philosophy of mind**

There has been a strong tradition in philosophy of mind of accounting for the relationship between the mental in the physical in terms of supervenience. There are many technical discussions of how exactly to understand the notion of supervenience, but the rough idea is that for the A-facts to supervene on the B-facts is for the B-facts to necessarily fix the A-facts. Plausibly the economic facts supervene on facts about the actions of economic agents:
the facts about people buying, selling, working, making contracts, etc. necessarily determine the facts about the inflation, stock market prices, the value of the Euro, etc. To a rough approximation: you couldn’t have two possible worlds perfectly alike in terms of the behaviour of economic agents, and yet different in terms of their economic facts.\textsuperscript{33}

However, there is now a broad consensus that we can’t define physicalism, or any other metaphysical view of fundamental reality, in terms of a merely modal relationship like supervenience.\textsuperscript{34} Counterexamples to a supervenience account of physicalism are raised by certain forms of emergentism. Recall that according to emergentism fundamental entities emerge at higher levels of complexity. 19\textsuperscript{th} century British emergentists, for example, were committed to properties at the chemical, biological, and mental levels which were wholly irreducible to, and could not have been predicted from knowledge of, the more basic physical facts.\textsuperscript{35}

Suppose mental emergentism is true, and suppose further that, as some of the British emergentists seemed to believe, the psycho-physical laws of nature in virtue of which the mental emerges from the physical are metaphysically necessary. Following Terry Horgan we can call the resulting view ‘Moorean emergentism.’\textsuperscript{36} Moorean emergentism entails that the physical facts necessarily fix the mental facts, for the laws in virtue of which the mental arises from the physical hold in all possible worlds. And yet this is clearly not a view according to which the mental is nothing over and above the physical; ex hypothesi the emergent properties are genuinely new fundamental features of the world.

These concerns about supervenience accounts would not go away if we added that the emergence must be intelligible, for example by their being an a priori entailment from the material facts to the mental facts. This is because, as we shall explore in chapter 6, emergence can sometimes be intelligible. Suppose matter partly consists of dispositional properties to produce consciousness, dispositions which remain latent most of the time but kick in at certain levels of complexity. If the essential nature of matter involves having the

\textsuperscript{34} See for example Horgan 1993, Melnyk 2003, Wilson 2005.
\textsuperscript{35} For examples of British emergentism, see Mill 1843, Broad 1925, Alexander 1920. For a good discussion of British emergentism see McLaughlin 1992.
\textsuperscript{36} Horgan 2006.
disposition to produce consciousness under condition C, then one could move a priori from knowing the nature of matter is in condition C to knowing that it will produce consciousness. It would not follow that consciousness is nothing over and the material facts.

Whilst physicalism cannot be completely accounted for in terms of supervenience, it is almost universally agreed that physicalism entails that all facts supervene on the facts. If the non-(narrowly) physical facts are nothing over and above the (narrowly) facts – if a world in which the fundamental facts obtain is already thereby a world in which all other facts obtain – then it’s hard to see how the non-fundamental facts could fail to be necessitated by the fundamental facts. Hence, I will assume that the supervenience of all facts on the narrowly physical facts is a necessary even though not sufficient condition for physicalism.37

Despite the broad consensus that supervenience accounts of physicalism are inadequate, there is little consensus in philosophy of mind as to what should replace it. In what she calls the ‘powers-based subset strategy’, Jessica Wilson explicates the physicalist claim that the mental is nothing over and above the physical in terms of the fact that the causal powers of the mental are a proper subset of the causal powers of the physical.38 UK women are a proper subset of UK people, which gives us a clear sense in which the totality of UK women is nothing extra to the totality of UK people. Similarly, if indeed mental causal powers are a subset of physical causal powers, then this gives us a clear sense in which mental causal powers are nothing extra to physical causal powers.

The problem with Wilson’s account is that it is too restrictive; it gives us no way of making sense of the nothing over and above relation holding between things which are not powers. Whilst some philosophers take the nature of each and every property to be given by its causal powers – a view I will argue against in chapter 6 – most philosophers believe in the existence of categorical properties, i.e. properties the nature of which cannot be entirely specified in causal terms. Wilson’s account offers us no way to make sense of the thesis that higher-level categorical mental properties are nothing over and above lower-level

---

37 Although there is a broad consensus behind the view that physicalism entails that all facts supervene on the physical facts, it is questioned in Montero 2013. Similarly, Leuenberger 2014 and Skiles Forthcoming argue against the thesis that grounded facts are necessitated by the facts which ground them. deRosset (2010) and Torgdon (2013b) defend this thesis.

categorical properties. It is also not clear how the view can make sense of the idea that higher-level objects are nothing over and above lower-level objects.

Robert Kirk explicates the physicalist conception of the relationship between the mental and the physical in terms of redescription: the mental facts are nothing over and above the physical facts iff the mental facts are ‘pure redescriptions’ of the physical facts. And a proposition P, is a pure redescription, relative to some base description B, only if the truth of P depends on nothing outside of the reality specified by B. The following would presumably be an example:

(A) Base description: Rod, Jane and Freddy are reveling.

Pure redescription (a): There is a party.

Intuitively pure redescriptions are just a different way of carving up the reality specified by the base description.

I feel the intuitions behind Kirk’s account. But the problem – put in terms of the above example – is that the account doesn’t tell us anything about the party itself. If the redescription <there is a party> is true, then then there is a party, out there in reality. What we want is an account of the fact that the party is nothing over and above the fact that there are people revelling. But Kirk tells us nothing about the party itself; he merely tells us something about a proposition which is about the party.

The only way I can think of to extend the account such that it accounts for the nothing over and aboveness of the party itself, is to add that the state of affairs of there being a party is grounded in the truth of the proposition <there is a party>. For we could then say that the fact that there is a party is grounded in the truth of <there is a party>, which is a pure redescription of the state of affairs of Rod, Jane and Freddy revelling (relative to the base description <Rod, Jane and Freddy are revelling at Jane’s>). But this gets things the wrong way round: propositions are representations of reality, and are true in virtue of correctly representing reality. The proposition <there is a party> is true in virtue of there being a party, and not vice versa. And once we accept this, it’s hard to see how claims about

39 Kirk 2013.
representations of the party can shed light on the nothing over and aboveness of the party itself (similar difficulties fact truthmaker accounts, discussed below).

Andrew Melnyk accounts for the physicalist conception of the relationship between the physical and the mental in terms of the nature of mental properties rather than mental concepts, and his account is very similar to the grounding by analysis model I outline above. I have disagreement with his view only in details. Melnyk’s account is focused on the notion of realisation, which in his 2003 book he defines as follows:

Token $x$ realizes token $y$ iff (i) $y$ is a token of some functional type, $F$, such that, necessarily, $F$ is tokened iff there is a token of some other type that meets condition, $C$; (ii) $x$ is a token of some type that in fact meets $C$; and (iii) the token of $F$ whose existence is logically guaranteed by the holding of condition (ii) $x$ is numerically identical with $y$.  

Melnyk’s use of the words ‘realisation’ and ‘functional’ is slightly misleading, as these words are generally used in relation to causally defined properties whereas Melnyk’s notion of a ‘functional type’ is tied to the meeting of a condition which may or may not be causal. A functional type is a ‘a higher-order type such that, necessarily, it is tokened iff there is a token of some or other lower-order type that plays some particular role or – more generally – meets some particular condition.’

In this first formulation, it was not explicit that the necessary connection between the token of the functional type and its token realiser obtains in virtue of the nature of the functional type. In the absence of this, it is not clear that Melnyk’s realization account of physicalism avoid the troubles faced by supervenience accounts of it. The Moorean emergentist might hold not only that mental properties are necessitated by physical properties, but also that there is no possible world in which mental properties exist uncaused, free floating as it were. On such a view mental properties would seem to be ‘realised’ by physical properties, going by the letter of Melnyk’s definition (we can take $C$ to be the condition causing mental

---

properties), even though this is clearly not a view according to which mental properties are nothing over and above physical properties.

In a forthcoming paper Melnyk is explicit about the connection to essence, in giving the following account of physical realisation:

1. \( M \) is a token of a mental state type with a certain higher-order essence; that is, \( m \) is a token of a mental state type \( M \) such that for a token of \( M \) to exist just is for there to exist a token of some (lower-order) state type such that tokens of that (lower-order) state type play role \( R_M \), the role distinctive of \( M \);
2. \( p \) is a token of physical state type \( P \) such that, necessarily, given the physical laws, tokens of \( P \) under physical circumstances \( C \) play role \( R_M \), and
3. the physical laws hold and physical circumstances \( C \) obtain.\(^{42}\)

The resemblance between Melnyk’s view and Dasgupta’s is striking, which is I think quite revealing about the absence of interaction between philosophy of mind and metaphysics on these foundational questions (N.B. Melnyk’s original exposition of this view was published at least twelve years before Dasgupta’s paper will be published).

However, note that the above is not a general notion of realisation, but a notion of physical realisation. We could of course return to the 2003 definition, altering it to make the connection to essence explicit. But condition (ii) of the 2003 definition states that the realizer in fact meets the condition required for the instantiation of the functional type. This is not sufficient to make the realised token nothing over and above the realiser, for it may be that the realiser token \( p \) satisfies the instantiation condition in virtue of contingent facts about \( p \), such as causal powers \( p \) has contingently in virtue of the laws of nature (in such a case the mere existence of \( p \) will fail to necessitate the realised token). This is why Melnyk ultimately defines physicalism not as the view that all tokens are realised by physical tokens, but that all tokens are physically realised in physical tokens (according to a notion of physical realisation similar to that given above), but we are thereby left without an adequate general notion of realization. I have no doubt that this tiny difficulty could be resolved, and I would probably be happy to embrace the resulting theory of fundamentality.

\(^{42}\) Melnyk forthcoming.
Accounts of fundamentality in metaphysics

Jonathan Schaffer, although a proponent of grounding, is inclined to think that relationships between levels of reality are mediated by brute metaphysical principles. It is perfectly coherent to suppose that fundamental particles never compose composite objects; and without brute composition laws, according to Schaffer, they never would. This kind of view does not satisfy the Free Lunch Constraint: if it takes brute metaphysical laws to get table facts out of particle facts, then the table facts are something over and above the particle facts. And I think we can do better. Although it is not strictly speaking a contradiction to say ‘There are people revelling but there is no party’, the fact that there are people revelling logically entails all that is essentially required for there to be a party. Similarly, although the particle facts do not logically entail that there are tables, it is plausible that they entail all that is required for there to be tables.43

One influential alternative to grounding conceptions of fundamentality, developed by John Heil amongst others, focuses on truthmaking.44 In Heil’s view, a great error in twentieth century metaphysics was the Quinean orthodoxy of reading off ontology from the entities quantified over in the sentences we take to be true.45 In this framework, avoiding an ontological commitment to Xs requires avoiding quantifying over Xs, or at least analysing sentences involving quantification over Xs into sentences not involving quantification over Xs. Contra this Quinean tradition, Heil thinks that ontology should focus not on the entities quantified over in our truths, but on the entities which serve as truthmakers for such truths. By holding for example that the truths about tables are made true by states of affairs involving particles being arranged in certain ways – arranged ‘table-wise’ as philosophers tend to say – we avoid an ontological commitment to tables.

43At least if we think of a ‘table’ as something that could be used as a table, such that a table come into existence by freak arrangement of particles even if there were no people around. For in this case it seems to me that all that’s required for there to be a table is for there to be a table-shaped region of space which satisfies a certain functional role, e.g. if a glass where to be put in such and such a location it would stay upright, etc. And it seems plausible in principle that this functional role could be specified in such a way that it would be logically entailed by the micro-level facts; even though it would probably be as a matter of fact impossible for any philosopher to actually state functional role with enough precision to show that. If we take it to be part of the definition of that it was manmade, then it’s much less clear that the particle facts could logically entail what is essentially required for there to be a table, but only because it’s much less clear that the particle facts could logically entail what is essentially required for there to be mentality.


45Originating in Quine 1948.
I think the truthmaking account faces difficulties similar to those facing Kirk’s account discussed above. Why should claims about representations of tables – e.g. claims about how they are made true – tell us anything about the metaphysical status of tables themselves? Proponents of the truthmaking approach often talk as though non-fundamental entities do not exist, saying for example that ‘in reality’ there are no tables only table-wise arrangements of particles. But if the view is that table-wise arrangements of particles make it true that there are tables, we seem on the face of it to have a contradiction. If it is true that there are tables then there are tables, which is inconsistent with there being no tables. And if tables do exist, in every sense in which their parts exist, it doesn’t seem to make sense to say that we avoid a commitment to them. If they exist, then we’re committed to them.

One way to qualify something like the truthmaking approach in order to avoid contradiction is to adopt what we might call metaphysical elitism. According to the metaphysical elitist, worldly entities are not equals; some are metaphysically privileged, participating in the metaphysically heavyweight structure of reality. Theodore Sider is a prominent proponent of this view. On his version the privileged structure of reality is captured in the ‘Book of the World’: the true and complete description involving only concepts which ‘carve nature at the joints.’ All other truths have ‘metaphysical truth conditions’ specified in the metaphysically privileged language (i.e. the language involving only terms which carve nature at the joints) and satisfied by metaphysically privileged truths. Thus contradiction is avoided: it is false in the metaphysically privileged language that ‘tables exist’, but it is true in a common or garden language like English that ‘tables exist’. Tables exist, but they are not part of the structure of reality.

Some will be suspicious of the use of primitive notions in metaphysics which don’t seem to have any analogue outside of metaphysics. As Sider himself confesses:

---

47 In the metaphysically privileged language the quantifier carves nature at the joints, and hence has a different meaning to the quantifier of an ordinary language like English.
'I know from bitter experience that philosophers are wary of this primitivism. Many times I have been asked (to murmuring general approval): “What on earth do you mean by ‘structure’?”'\(^{48}\)

In response to this, Sider tries to make the notion ‘earn its keep’ by demonstrating its theoretical utility. However, Jonathan Schaffer has persuasively argued that Sider fails to do this: Sider invests in a non-comparative notion of privilege – concepts either carve nature at the joints or they don’t – whereas it is a comparative notion of privilege which seems to be needed for the theoretical uses to which Sider wants to put this notion, for example, in accounting for laws of nature and the determinacy of reference.\(^{49}\)

Moreover, privileged structure seems redundant in meeting the Free Lunch Constraint. Suppose, as Sider does, that the English sentence ‘There are tables’ is true iff there are particles arranged table-wise. When speaking English, we can say all the things we would say if we adopted a grounding account of fundamentality: there are tables, and there are tables because there are particles arranged table-wise. And I can see no reason not to take the explanatory relationships expressed by the word ‘because’ in the last sentence to be a grounding relation. Grounding relations obtain between existent things, but – so long as we are speaking English – tables do ‘exist’.

Even if it is true that tables don’t ‘Exist\(_H\)’, i.e. in the metaphysically heavyweight sense expressed by existence terms which carve nature at the joints, they still ‘exist\(_L\)’, i.e. exist in the metaphysically lightweight sense expressed by existence terms in ordinary English. Hence we can talk about essential truths concerning tables, understood as truths concerning what it is for tables to exist\(_L\), and explain the ontological free lunch in terms of the grounding by analysis model. Further claims about what is or isn’t true in the metaphysically privileged language are surplus to requirements.

I share Sider’s intuition that some facts are metaphysically privileged, that the world has a ‘deep structure’. But we can account for this sense of being privileged in terms of grounding: the privileged facts are those substantive facts which are not grounded in other substantive facts. This order of explanation seems to me more satisfying that starting off

\(^{48}\) Sider 2012: 9.
\(^{49}\) Schaffer 2014.
with a primitive notion of ‘privileged structure’ which has no analogue outside of metaphysics.\textsuperscript{50}

\textsuperscript{50} To be honest, I go back and forth on this. In Goff Forthcoming, I try to show that how we can make sense of notion of privileged existence being present in ordinary language and thought. At any rate, I don’t think of the Sider-type view as inconsistent with a grounding conception of fundamentality; it just adds a further level of analysis.